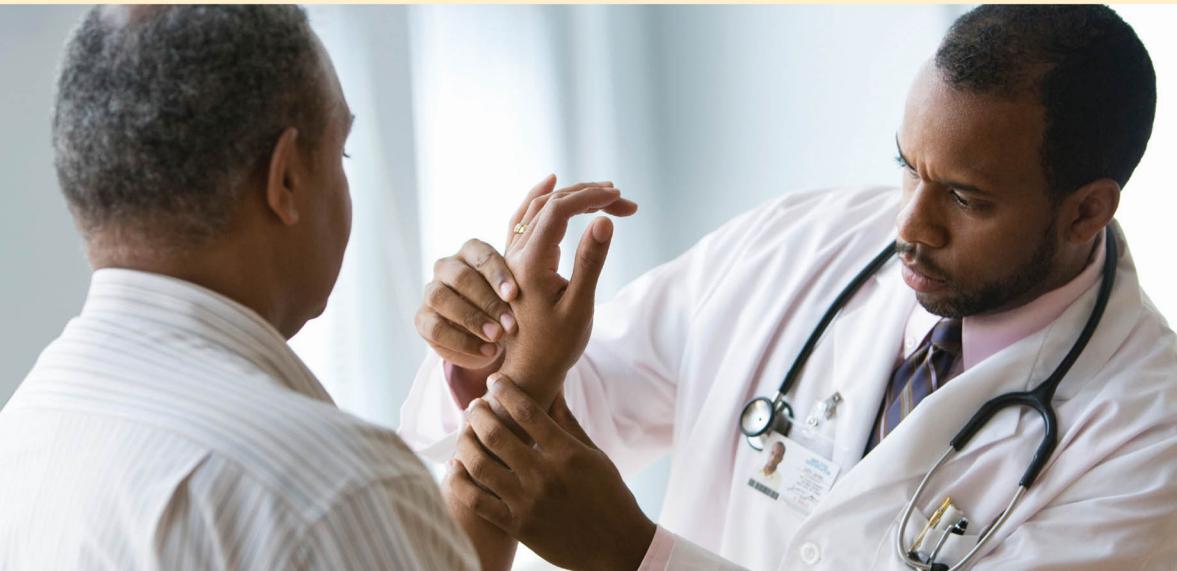


THE PRAEGER HANDBOOK OF

# OCCUPATIONAL AND ENVIRONMENTAL MEDICINE



TEE L. GUIDOTTI, M.D., M.P.H.

**THE PRAEGER  
HANDBOOK OF  
OCCUPATIONAL AND  
ENVIRONMENTAL  
MEDICINE**

*This page intentionally left blank*

# **THE PRAEGER HANDBOOK OF OCCUPATIONAL AND ENVIRONMENTAL MEDICINE**

---

**VOLUME**

**I**      Principles

Tee L. Guidotti, MD, MPH



AN IMPRINT OF ABC-CLIO, LLC  
Santa Barbara, California • Denver, Colorado • Oxford, England

## **Copyright Acknowledgments**

Material from two chapters adapted from Tee L. Guidotti. "Taking the Occupational History." *Ann Intern Med.* 1983; 199-4: 641-651.

Copyright © 2010 by Tee L. Guidotti

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, except for the inclusion of brief quotations in a review, without prior permission in writing from the publisher.

## **Library of Congress Cataloging-in-Publication Data**

Guidotti, Tee L.

The Praeger handbook of occupational and environmental medicine / Tee L. Guidotti.

p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-313-35999-6 (set hard copy : alk. paper)—ISBN 978-0-313-36001-5 (vol 1 hard copy : alk. paper)—ISBN 978-0-313-36003-9 (vol 2 hard copy : alk. paper)—ISBN 978-0-313-38204-8 (vol 3 hard copy : alk. paper)—ISBN 978-0-313-36000-8 (set ebook)—ISBN 978-0-313-36002-2 (vol 1 ebook)—ISBN 978-0-313-36004-6 (vol 2 ebook)—ISBN 978-0-313-38205-5 (vol 3 ebook)

1. Medicine, Industrial—Handbooks, manuals, etc. 2. Environmental health—Handbooks, manuals, etc. 3. Medicine, Industrial—Handbooks, manuals, etc. I. Title. II. Title: Praeger handbook of occupational and environmental medicine.

[DNLM: 1. Occupational Medicine. 2. Environmental Health. 3. Environmental Medicine. 4. Occupational Health. WA 400 G948p 2009]

RC963.32.G85 2010

616.9'803—dc22 2009046237

ISBN: 978-0-313-35999-6

EISBN: 978-0-313-36000-8

14 13 12 11 10 1 2 3 4 5

This book is also available on the World Wide Web as an eBook.

Visit [www.abc-clio.com](http://www.abc-clio.com) for details.

Praeger

An Imprint of ABC-CLIO, LLC

ABC-CLIO, LLC

130 Cremona Drive, P.O. Box 1911

Santa Barbara, California 93116-1911

This book is printed on acid-free paper 

Manufactured in the United States of America

# CONTENTS

## VOLUME 1: PRINCIPLES

<b>Preface</b>	<b>vii</b>
1. Occupational and Environmental Medicine	3
2. Toxicology	55
3. Epidemiology	201
4. Assessment of Exposure	295
5. Monitoring, Surveillance, and Screening	327
6. Safety and Injury Prevention	357
7. Risk Science and Risk Communication Skills	379
8. Prevention Science	459

## VOLUME 2: CENTRAL ISSUES

9. Physical Hazards and Conditions	505
10. Common Chemical Hazards	569
11. Buildings	689
12. Environmental Health	739
13. Stress and Psychosocial Factors at Work	809
14. Hazard Control	845
15. Clinical Evaluation	889
16. Occupational Injuries and Musculoskeletal Disorders	949

## VOLUME 3: PRACTICE INSIGHTS

17. Diseases and Health Protection	1013
18. Capacity for Work	1131

## **Contents**

19. Wellness and Health Promotion	1201
20. Occupational Health Services	1245
21. Workers' Compensation	1315
22. Emergency Management	1357
23. Medicolegal Services	1389
24. Regulatory Agencies	1443
25. Society, Critical Science, and Ethics	1481
26. Global Occupational and Environmental Health	1537
<b>Conclusion</b>	<b>1585</b>
<b>The Essential Library for the Occupational and Environmental Physician</b>	<b>1597</b>
<b>Index</b>	<b>1599</b>

# PREFACE

This set is an introduction to one of the most dynamic and challenging fields of practice in medicine. It is also a rethinking of the field, intended to put the pieces together in a certain way that balances theory and practice, occupational and environmental, traditional and cutting-edge.

The set is intended for:

- Physicians who are entering the field midway through their practice and wish to prepare themselves systematically for a new career
- Physicians who intend to specialize in occupational and environmental medicine
- Physicians currently practicing occupational medicine who wish to rethink their practice
- Health professionals, scholars, and students who wish to understand the structure, history, and content of occupational and environmental medicine
- Health professionals engaged in occupational and environmental medicine who want an overview and an integrative framework that puts the pieces together

These volumes were designed for the physician who is already well trained in clinical medicine and does not need an extensive discussion

on medical or surgical treatment of common disorders. Although some textbooks in occupational and environmental medicine do emphasize treatment, these volumes do not. Clinical guidance is readily available elsewhere. Instead, these volumes are designed provide the practicing occupational physician with what he or she needs to know that a good physician may not know already, and to put occupational and environmental medicine into a coherent perspective for the OEM physician who is already experienced. Where they conspicuously fall short is in failing to unify occupational and environmental medicine seamlessly. Although the sister fields overlap substantially and are clearly merging, occupational medicine is still practiced mostly in a self-contained health care system; and environmental medicine, as conceived of here, is still primarily the medical face of a field of public health, not clinical medicine. It may take another generation for the two to become one fully unified field, but in the meantime they are treated in this set of volumes as poles defining opposite ends of a single field and are frequently referred to separately.

The author and the publisher welcome corrections and suggestions for improvement of the text. However, it is not practical to engage in long discussions about content. The reader is advised that although facts have been checked, it is not practical to insert citations in what is intended to be a flowing narrative. The use of citations is becoming increasingly less helpful to readers because of the ease of Internet searches, which generally take much less time these days than tracking down a book from a reference and are likely to be more up to date. Authoritative documentation should always come from original sources, not secondary sources such as textbooks.

Finally, this set is dedicated to the memory of Clifford Rodney May, LMSSA, MSc, DIH, FFOM, FFPHP, FRCP(C), FACOEM. An extraordinary occupational physician and role model, Rodney left his mark on the field in Canada, the United Kingdom, and the United States. No doubt he is also still remembered in some corner of Singapore.

# **VOLUME I:**

## **Principles**

*This page intentionally left blank*

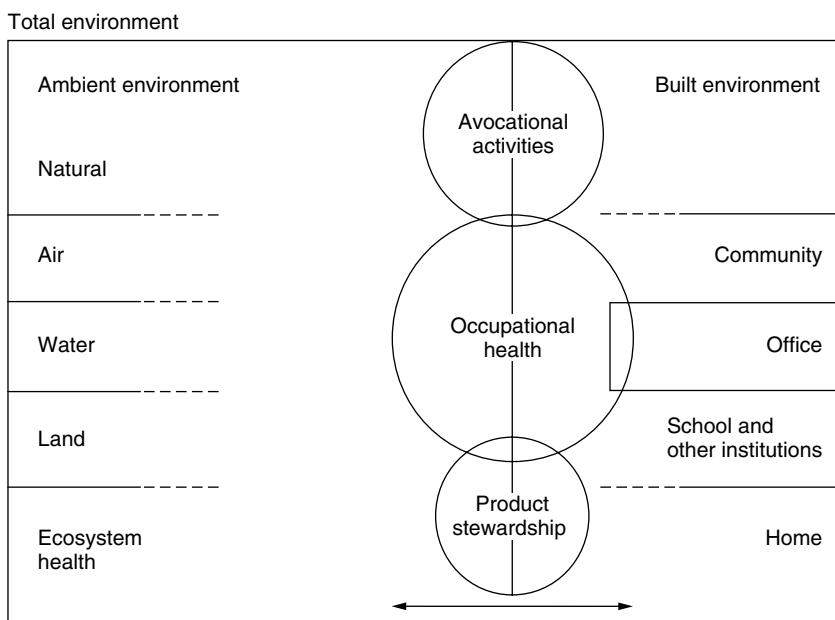
# **1 OCCUPATIONAL AND ENVIRONMENTAL MEDICINE**

Occupational and environmental medicine (OEM) is a unitary field of medical specialization and an integrating field of practice that unites a preexisting and quite old medical field, occupational medicine, and the closely related and equally old but now rapidly emerging field of environmental medicine, which has roots shared with public health.

OEM comprises the role of the physician in the broad field of environmental health, including and with special emphasis on that part that deals with workplace environmental health issues, occupational health. The field also includes closely related domains such as exposures in the home and schools, health issues associated with product stewardship, avocation (such as hobbies), and, perhaps with less confidence, health issues associated with large-scale environmental issues such as climate and ecosystem change (see Figure 1.1). Virtually everything that a physician in OEM does can be mapped somewhere within this conceptual landscape.

The total environment is divided, for purposes of environmental health sciences, into the “ambient” environment (outdoors) and the “built” environment (what human beings have created, mostly indoors). The ambient environment can be divided (the divisions

are just suggested) into the natural environment (which itself presents many hazards and health challenges), media (air, water, and land, which for the purposes of this schema includes food grown on it), and ecosystem health (including climate change, stratospheric ozone depletion, and ecosystem destabilization on a smaller scale). The built environment ranges in scale from communities (cities to rural districts) and all their attendant infrastructure (markets, roads, energy grids, bridges, dykes and levees, and so forth), to workplaces with and without exceptional risk, schools and other institutions, and homes (whether multiple dwellings or individual houses). Occupational health is traditionally most concerned with workplaces associated with exceptional risk (factories, workshops, small enterprises), but also deals with workplaces where workers are not at exceptional risk, such as offices. Occupational health overlaps with environmental health concerns in both the ambient and built



**Figure 1.1.** Environmental health and its subfields and how they fit together.

environments (because some workers work outdoors and most work indoors), with avocational activities (hobbies and activities that are not remunerative but share many of the same hazards), and with product stewardship (consumer safety and the management of risk). Perhaps subtly, at the bottom of Figure 1.1 there is no division between the ambient and built environments, because each affects the other.

Terminology can be confusing in OEM, as in any field in rapid transition. Terms and phrases will be used as precisely as possible in this book. The name “occupational and environmental medicine” accurately describes the broad interests of OEM practitioners in the relationship between human health and influences external to the body, those arising from the environment, whether natural, outdoors (ambient), constructed (built), or one of the most important built environments, the workplace. OEM, while it is becoming a unitary field, still has two poles in practice and sometimes it is necessary to talk about them separately. In this book, “occupational medicine,” when it is not used to refer to the recognized medical specialty as described below, means medicine and health protection applied to the workplace. “Environmental medicine,” when not otherwise specified, means medicine and health protection applied to the natural, ambient or built environment outside the workplace.

In the United States and Canada, as in many countries, there is a recognized medical specialty called simply “occupational medicine” that, in both countries, incorporates content that is broadly recognized as environmental medicine. This specialty is simply OEM on the level of specialized medical practice. Many physicians practice in the field of occupational medicine without being specialists, and many specialists, the author among them, would prefer that the specialty be called “occupational and environmental medicine” officially.

In this book, OEM (usually so abbreviated) will refer to the broad, unified field of practice. Occupational medicine and environmental medicine will be treated separately only when there is reason to do so, as in this chapter. When there is a need to discuss the medical specialty of occupational medicine in the context of specialty affairs, this will always be made explicit.

The integration of OEM is most complete in corporate practice and in academics (discussed in more detail below), less complete in government, because of the fragmentation of responsibility among agencies, and least complete in private practice. Practice in the corporate private sector has become a unitary field, because responsibility for environmental health usually comes with senior positions in OEM. In the public sector, OEM physicians in government often move among regulatory and other government agencies in the course of their careers but are usually limited by the mandates of their agencies while they hold a given position; even so, some agencies have dual responsibilities (such as the EPA). Other than litigation support, there is no obvious business or payment model (to compare with workers' compensation) for environmental medicine in private practice. Insurance companies usually do not recognize, approve, or reimburse for consultations for environmental disease or exposure. In real terms, occupational medicine is a mature field of practice and environmental medicine is not, at least not yet, in private practice. So the OEM practitioner usually supports the practice with occupational medicine and does environmental medicine as the opportunity arises or out of commitment.

OEM is already well integrated as an academic field because both the occupational and environmental sides draw from the same knowledge base—emphasizing toxicology—and methodology—emphasizing epidemiology. As seen from academe, the two sides of OEM differ mostly in scale, exposure levels, and characteristics of the population at risk (the general population or the working population, which is generally healthier). Properly speaking, occupational medicine, focused as it is on the work environment, should logically be subsumed under the broader field of environmental medicine, as in Figure 1.1. Many OEM physicians in academic institutions do research on public health aspects of environmental hazards and teach environmental health, and this is reflected in the name of some academic departments.

In practice, occupational medicine and environmental medicine do operate on two distinct but overlapping planes. The occupational

medicine side of OEM, on the one hand, embraces a vast range of medical and public health issues but is, at its heart, about the relationship between work, capacity, and the health of individuals and relatively small groups defined by work and community. This may take the form of exposure to chemical and physical hazards in the workplace and the effect of such exposure on human health. It may also take the form of the limitations and reduced capacity for work and daily living imposed by health conditions. Historically, occupational medicine has emphasized the physical and social characteristics of the workplace as a specific environment and as a microcosm of society. It has a history, a clear identity, and has been shaped by social trends, economic forces, and humanistic values. The web of job requirements and responsibilities, regulations, credentials, regulatory agencies, technology, and professional licensure requirements in which occupational and environmental medicine is embedded also tends to keep OEM physicians focused on the occupational side.

Environmental medicine is still coming into focus as a domain of medical inquiry and is not established in practice. The environmental medicine side of OEM embraces an even more vast but diffuse spectrum of issues, and at its heart is the relationship between living, the surrounding world, and the health of both individuals and entire communities or larger populations. It has never been a recognized specialty. Nonetheless, environmental medicine has deep roots itself and a long history intertwined with public health. During the great era of discovery in public health in the early and middle twentieth century, medical aspects of air quality and toxicology were usually investigated by occupational physicians and those of poor water quality by preventive medicine and public health practitioners. Preventive medicine specialists are historically associated with the administration of public health departments, provision of preventive medicine services to the community, and broad issues of environmental health policy. OEM physicians working in environmental health are usually focused on more specific issues related to the evaluation of health risks, product stewardship, global public health (because of travel and business abroad), health

protection in institutional facilities and regulatory compliance. Both sides are concerned with health promotion and disease prevention.

This textbook reluctantly takes the convenient but potentially divisive expedient of treating environmental medicine separately when the discussion requires, as is done later in this chapter and in Chapter 12.

## **OEM AS A FIELD OF PRACTICE**

Occupational and environmental medicine is an unusual field of medical practice. There is virtually no inpatient element of practice, which makes the field almost invisible to the rest of medicine and to medical students. Much of the medical knowledge inherent in OEM lies outside traditional clinical practice, especially with regard to issues of population health, compensation, and medicolegal services. The field blends population health management with individual patient care and has done so, consciously, for at least a century. New drugs and devices hardly matter in the sense that they change practice very little, because practice of OEM is driven primarily by the economy and changes in technology.

Most new practitioners enter the field in midcareer, after they have practiced in other specialties, and for that very reason most practitioners in the field cannot afford to take time out for the specialty training programs and so do not become board-certified despite their competence. They prepare themselves in other ways, including short courses and continuing education. Almost uniquely in medicine, however, this “lateral entry” is not a shortcoming (although at times it has been perceived as such). It keeps the field alive by meeting demand in the community and by attracting seasoned, mature clinicians with life experience. Many of these lateral entrants excel in the field and provide valuable service to communities.

The specialty is recognized by the American Board of Medical Specialties under the name of “occupational medicine” and is certified in the United States through the American Board of Preventive Medicine. The principal organization in the field in the United States

is the American College of Occupational and Environmental Medicine (ACOEM), which, with 5,000 members, is the largest medical specialty organization in the field worldwide. At least 15,000 physicians practice OEM in the United States, mostly providing care and evaluation for work-related injuries and illnesses. Most are board certified in some specialty of medicine but not necessarily in occupational medicine.

In Canada, the fellowship of the Royal College of Physicians and Surgeons of Canada also recognizes the specialty using the name “occupational medicine,” though it is organized as a subspecialty within internal medicine. There is a second certification body, the Canadian Board of Occupational Medicine, for physicians who are not formally trained. In Canada, the principal organization for the field is the Occupational and Environmental Medical Association of Canada, with about 350 members. There are many other strong national organizations of occupational physicians, such as the Society of Occupational Medicine in the United Kingdom. Most of them incorporate environmental medicine to some degree but emphasize occupational medicine.

There is a strong network of international organizations led by the International Commission on Occupational Health (ICOH), which has about 2,000 members worldwide. ICOH represents all occupational health professions and disciplines among its members, not only physicians, although they are heavily represented. ICOH has a variety of partnerships with other international organizations dealing with specific occupational health professions and with social insurance and workers’ compensation. ICOH is recognized as a non-governmental organization by the United Nations and partners on projects with the World Health Organization and the International Labour Organization.

OEM is greatly limited in North America by longstanding constraints on physician supply. There are only approximately 3,000 physicians boarded in occupational medicine in the United States, and only around 70 new certificants take their boards every year from the American Board of Preventive Medicine (ABPM). Almost all of

them already have board certification in another specialty, most commonly internal medicine, family medicine, or emergency medicine. To qualify, they must complete a three-year residency program that includes one year of clinical training (although virtually all programs require completion of three years of clinical training prior to entry), one year of didactic instruction leading to the master's of public health degree (MPH) or its equivalent, and one year of supervised practicum experience in occupational health settings. Many leaders in the field think that these nominal requirements do a disservice to the specialty, and there is a movement to replace them with a more honest five-year requirement—with at least three years of clinical training—and to replace the residency with a fellowship model. (See the Conclusion, in Volume 3, for further discussion.)

Training programs in OEM share content with other specialties, particularly in primary care and those specialized areas important for occupational health: clinical practice (management of musculoskeletal disorders, pulmonary, dermatology, neurology), epidemiology, toxicology, prevention/health promotion, medicolegal analysis, and population medicine, by which is meant addressing the health of groups of people through organized programs and risk management. The field has unique content, including knowledge of occupational and environmental exposures, specialized diagnosis for diseases of importance (such as the pneumoconioses), disability evaluation, accommodation and fitness for duty, environmental health, management science (not health services management but the corporate variety), and emergency preparedness (not emergency medicine but the management of mitigation, preparation, response, and recovery at the enterprise level).

For a relatively small field, occupational and environmental medicine has a remarkably strong research and training infrastructure. Top journals include the *Journal of Occupational and Environmental Medicine* (the journal of ACOEM), the *American Journal of Industrial Medicine*, *Occupational and Environmental Medicine*, *Occupational Medicine* (often called the “yellow journal”), *Environmental Health Perspectives* (the journal of the National Institute of Environmental Health Sciences, or NIEHS), and the *Scandinavian Journal of Work, Environment and Health*.

Historically, the most influential journal in OEM used to be *Archives of Environmental and Occupational Health*, which, although much reduced today, was the leading journal for occupational medicine, occupational (industrial) hygiene, and air pollution studies for most of the twentieth century. *Archives* has been published under many names over the years and absorbed several earlier journals. The journal is currently making a comeback.

The lead federal agencies that support OEM are the National Institute of Occupational Safety and Health (NIOSH, which funds academic training and research centers of excellence), the Agency for Toxic Substances and Disease Registry (ATSDR, which is also, with NIOSH, a part of the Centers for Disease Control, or CDC), and the National Institute of Environmental Health Sciences (NIEHS).

The daily work of physicians practicing OEM is highly individualized, much more than other medical specialty areas. Most practitioners provide acute care for work-related injuries and illness and, of course, deal every day (in the United States) with the Occupational Safety and Health Administration (OSHA) and its state counterparts and with numerous agencies and carriers in the world of workers' compensation. However, there is also considerable diversity, particularly for board-certified specialists: chronic disease management, work-related injury and illness care, workers' compensation, medicolegal services, environmental medicine, managed care, health promotion and disease prevention, and research. Also, there are an unusual number of "niche" areas in the field for practitioners with specialized qualifications, including drug testing, conducting specialized patient assessments ("independent medical evaluations"), and applying the regulations of the U.S. Department of Transportation (the health of drivers and other transport workers constitutes an important part of practice).

## **HISTORY OF OCCUPATIONAL MEDICINE**

As the narrower field with a tradition of medical service, the history of occupational medicine is better defined and more restricted

in scope than that of environmental medicine. The history of environmental medicine, which dates to Hippocrates' *On Airs, Waters, and Places*, is largely the history of environmental health and public health as a whole and is beyond the scope of this book.

The story of occupational medicine is less familiar and needs to be told because events from long ago still have relevance today to attitudes and practice. The history of occupational medicine has a wealth of fine detail and interesting anecdotes, but the big picture is most important to understand.

How the field came about and why it is in the position it is in today should be known by all OEM practitioners. It explains a great deal about the specialty, trends in the field as a whole, and the status of occupational physicians in workers' compensation and in organized medicine. For example, it helps to know that occupational medicine tends to swing through extreme cycles of alternating growth and struggle, with a period of about thirty years. Occupational medicine enjoyed peaks of interest and enthusiasm in the United States around 1910 (associated with the Progressive Era in American politics), 1940 (associated with the Second World War), and 1970 (associated with the Occupational Health and Safety Act). Almost on schedule, the field is currently on the ascendancy again, coming out of a prolonged period of stasis, neglect, and political weakness. These thirty-year cycles have been the bane of the specialty and have placed it at a severe disadvantage compared to other fields of medicine.

Prior to about 1950, occupational medicine was called "industrial medicine," a name that persisted in common use for another twenty years and came to be considered pejorative. The transition to "occupational" came in part because the field sought to get away from a narrow focus on the employer or industry and to focus more on occupation and the needs of workers. That pull between service to the industry and employer and service to the worker has been a source of tension for at least 200 years. At times it has seemed to be the defining feature of the field (it is not), and at other times (such as now) it has fallen into perspective as one of many professional challenges. This

tension has made occupational medicine acutely aware of its history and also of its need for a strong ethical framework.

History matters deeply in medicine, and nowhere more than in occupational medicine. History defines the field, tells its story, explains how it got the way it is, and above all gives the field a claim on real heroes. Heroes, in turn, are important because they provide occupational medicine, as all fields, with role models of good practice, integrity, and myths, in the positive sense of founding beliefs and stories that inculcate values. When one refers to Alice Hamilton, for example, it serves as a kind of shorthand for the values she held and her legacy. History also provides warnings and cautionary examples, reminding the OEM in the present to be wary of assumptions about risk and to put the health and safety of workers first.

There have been at least two positive historical traditions in occupational medicine, one oriented toward the worker and valuing effective health protection mostly as a human right and the other oriented toward industry and valuing effective health protection mostly as an economic investment and a common good. The former reached its peak in the Progressive movement in American history in the early twentieth century and in the influence of Irving Selikoff in the late twentieth century. The latter, emphasizing corporate responsibility and the power of large organizations to achieve health gains, has been sustained mostly by enlightened corporate medical directors who enjoyed the support of employers with progressive policies.

There has also been an ugly side to the field, such as when physicians have compromised or disregarded the health of workers or pursued policies favorable to the employer. Assessing historical accountability and the behavior of the field's practitioners has created an ongoing tension within the field. The history of occupational medicine in the twentieth century is often characterized as a conflict between labor and industry in which occupational physicians were recruited or seduced into taking the employer's side, except for a few brave champions who resisted the trend from outside the mainstream. However, it is fair to question this simplistic, revisionist history. The closer one looks at history, the more it appears that the

asbestos apologists and other sellouts, not the “good guys,” were the outliers in the long history of occupational medicine.

## Early History

Occupational medicine may be as old as medical practice itself. Curers, healing practitioners, shamans, herbalists, and bone-setters of the past must have treated many injuries that arose from the mammoth hunt and later treated many diseases that arose from working conditions. Medical and surgical care for injured workers was known to the ancients and seems to have been taken for granted as a normal part of practice. The ancient Egyptians employed physicians to care for ill and injured workers in the crews engaged in large building projects, such as temples, pyramids, and the Sphinx. Some records of their health problems and sickness absence days have survived.

Occupational diseases are only occasionally recorded in the ancient Greek and Roman medical literature. One reason for the lack of attention to the problem by ancient physicians is that most Western societies of the time had deep social class divisions. In ancient times, the most hazardous occupations, such as mining, were usually performed by slaves or condemned prisoners, and both groups were expendable. Workers and the lower-status physicians (often slaves themselves) who took care of them did not write books. It is known that miners used crude respirators made of pig bladders in an effort to reduce the amount of dust they inhaled, meaning that there was some understanding of prevention.

During the Middle Ages in Europe, at least toward the end, there was much activity related to local occupational health issues. Physicians and barber-surgeons were present in at least some mining camps and preindustrial economic communities. Some problems were well recognized. In particular, the disorders associated with mining and their prevention by ventilation were documented in treatises on mining written by Georgius Agricola (Georg Bauer) and by Ulrich Ellenbogen before 1600. Rickets, malnutrition, and lung diseases (which would have variously included tuberculosis, silicosis,

and lung cancer induced by radon daughters) experienced by silver miners in Bohemia resulted in very short life spans and physical deformities (particularly rickets and hunched back due to posture and possibly tuberculosis of the spine, Pott's disease), together with stunted growth due to malnutrition, were especially well known and recounted in central European legends, one of which became the story "Snow White and the Seven Dwarfs."

Health protection in the Middle Ages depended on traditional beliefs and superstitions, but some of the methods worked, to a greater or lesser degree. One of the more effective means of protection was designed to protect the physician himself (they were always men at that time). During plague times, physicians in the Middle Ages wore long-nosed masks filled with herbs and fragrant flowers in order to counteract the "bad airs" (the origin of the word "malaria," through Italian) or "miasmas," which were currents or drafts of contaminated air (Figure 1.2). The mask, together with heavy black velvet robes and an iron—rather than the usual wood—staff for pointing at and touching



**Figure 1.2.** The mask of "il dottore della peste," a reproduction of the masks worn by physicians attending victims of the plague. The character and its souvenir masks are still popular and figure prominently in the Carnevale of Venice.

patients, were thought to protect the physician from the miasmas that were thought to carry the disease. In reality, the mask mostly protected against the stench but kept the physician's face covered and would have prevented transmission of pneumonic plague. The costume also protected the wearer against fleas, the vectors of bubonic plague. As a consequence, mortality from plague among physicians was surprisingly low for a profession with such high exposure. On the other hand, mortality was extremely high among nuns, who served as nurses and actually cared for the ill. The figure of a physician wearing this costume became a recurring comic character in *commedia dell'arte*. Masks representing the unique beak-like image remain popular during Carnevale in Venice.

The towering but deeply flawed central European alchemist and mystic Theophrastus Philippus Aureolus Bombastus von Hohenheim, better known as Paracelsus (1493–1541), was a transformative figure in science and medicine. He led alchemy into the beginnings of modern chemistry and began the systematic, scientific study of toxicology. The story of this complicated, brilliant, and thoroughly unpleasant man will be briefly described in Chapter 2. Much of his work pertained, directly or indirectly, to occupational toxicology as it was understood at the time and especially the diseases of miners. He stood between superstition and science, with a foot in both camps, not only in occupational medicine and toxicology but in Western material philosophy as a whole.

### **The Age of Ramazzini, “The Father of Occupational Medicine”**

The birth of occupational medicine as a systematic, integrated medical field dates to the publication in 1700 of Bernardino Ramazzini's masterpiece *De Morbis Artificum Diatriba* (*Diseases of Workers*) (Figure 1.3). By careful observation and laborious verification by experience and questioning, Ramazzini accumulated in one text the most relevant facts about occupational disease extant in his era. He came to insights about mechanisms and opportunities for



**Figure 1.3** Bernardino Ramazzini, the “Father of Occupational Medicine.” (Source: [http://www.euuzlet.hu/galeria/2006\\_2\\_nagyedev.html](http://www.euuzlet.hu/galeria/2006_2_nagyedev.html))

exposure that are completely modern and that unfortunately still apply today, especially in the trades. He applied a disciplined mind to his observations and formulated astute conclusions about the social context of employment in his time, the nature of ethnic stereotyping, access to healthcare, and treatment outcomes.

Few physicians have had as much influence beyond their lifetime as Ramazzini (1663–1714). Born into a privileged family in Carpi, in the duchy of Modena in the territory of the Republic of Venice, he studied medicine under the best teachers in Parma and started a small-town practice in central Italy. He became ill with malaria and was forced to restrict his activities. His erudition and insight were slowly recognized and he became a founding member of a new medical school in Modena. There he became involved in research and engaged in some of the major medical controversies of the day. He was a close friend and colleague of other important medical figures

of the day such as Giovanni Morgagni and Marcello Malpighi, and of scientists such as Gottfried Leibniz. In 1690 and 1694, he published two major works on public health and disease risk in Modena and surrounding villages, which established him as an expert in environmental medicine. After many years of observation, he wrote his seminal text, following a series of lectures on the diseases of workers that he started in 1690. This was the first comprehensive work that gave shape, science, and even a sense of style (his Latin was considered highly stylish and erudite) to what became known as occupational medicine. On the basis of these works, he was recruited in 1700 to the great medical school at the University of Padua and became even more celebrated as a superb teacher and scientist.

Modena and Padua were territories under the political and economic control of Venice at the time, when the island republic, *La Serenissima*, was a world economic power. In certain technologies, such as glass, it was the dominant technological innovator worldwide. Venice was also the most important port in southeastern Europe, a relatively open society, and a center for publishing and book distribution. Padua was one of the greatest centers of learning in Europe. Venice and its possessions were therefore at the center of a vast network of communication, scholarship, and innovation in practical matters as well as abstract knowledge.

Ramazzini was a role model of the physician as scholar, scientist, and humanist. He represents in medicine the twilight years of a declining scholarly tradition that was cosmopolitan rather than national or local, European rather than Venetian or even Italian. Later authors, such as Charles Turner Thackrah, Benjamin McCready, and Alice Hamilton (described later in this chapter), wrote their works with an emphasis on conditions in their own countries. Ramazzini concentrated on the occupations themselves, their hazards, their social context, the conditions of life experienced by their practitioners, and the consequences of poverty. He took it for granted that his observations would be useful elsewhere, that conditions in his day were not unique, and that the same problems would be found elsewhere.

It is still common to hear references to Ramazzini when occupational medicine is discussed. His admonition to physicians always to ask the patient “What occupation do you follow?” is quoted often in the literature of occupational medicine. Another common reference is to “the three-legged stool,” which refers to Ramazzini’s admonition that the physician should listen to the patient attentively even if the patient’s home is so humble that there is only a stool to sit on. His descriptions of occupational disorders are often quoted in discussions of modern-day problems because they are in many cases still relevant. The paradox of Ramazzini’s contemporary relevance has nothing to do with prescience and everything to do with a lack of social progress since his time. The trades Ramazzini observed were all old by the time he observed them. Given his perceptive eye and inquiring mind, it is to be expected that once Ramazzini began his investigations he would see and describe in his work concepts and associations that are now considered to be core knowledge in occupational health, and a mind as perceptive as his would certainly generalize to basic principles.

Even his descriptions of the social order and the fatalism associated with employment (and the desperation of unemployment) are still relevant, particularly in less-developed societies. After three centuries, Ramazzini’s work is entirely too relevant to the situation today.

Sadly, Ramazzini’s work and his otherwise almost entirely progressive thinking are marred by virulent anti-Semitism. It is a sad reminder that even great and admirable minds may reflect the prejudice of their times.

## **The Industrial Revolution**

The Industrial Revolution began in England and Scotland in the latter part of the eighteenth century and proceeded unevenly first in Europe and later in North America through the nineteenth century. It was characterized by technological innovation and the introduction of new factories, but it also increased agricultural productivity and caused massive migrations of people, which in England depopulated

the countryside and brought a new labor supply to rapidly growing cities and industrial towns. Rural residents accustomed to the cycles of the growing season and informal practices of credit, payment in kind, and barter had to adjust to a new economy based on hourly wages and regular attendance at work. The organization of work changed dramatically, from the predominance of small workshops based on crafts, often in the home, to mass production in large factories or piecework in the home, with assigned quotas.

Although his life was shortened by tuberculosis, Charles Turner Thackrah (1795–1833) had a profound impact on occupational medicine in the English-speaking world (Figure 1.4). He was highly trained for his time but chose to practice in Leeds, a provincial city undergoing great transformation because of the Industrial Revolution. (He can be compared with the character Dr. Lydgate in George Eliot's *Middlemarch*.) He became interested in occupational health and in 1831 published the first work on occupational medicine in English: *The Effects of the Principal Arts, Trades, and Professions, and of Civic States and Habits of Living, on Health and Longevity: with Suggestions for the Removal of Many of the Agents which Produce Disease and Shorten the Duration of Life*. It was a masterpiece, noteworthy for excellent clinical descriptions of lead toxicity and other disorders, but also thoroughly modern in concept, emphasizing prevention and the control of exposure. His book established that occupational disease was a serious problem in England and laid the groundwork for later reforms.

Louis-René Villermé (1782–1863), “Le premier observateur de la santé au travail,” was the leading advocate of occupational medicine in France. He was part of the public health and political reform movement that swept France in the nineteenth century as the children of the French Revolution, full of high-minded expectations, grew up to encounter the reality of conditions such as those described in Victor Hugo's *Les Misérables*.

In the early nineteenth century, American medicine still looked to Britain, first England and later Scotland, for medical knowledge and standards of practice, even though Austria, France, and Germany were



**Figure 1.4.** Charles Turner Thackrah, as represented on a commemorative souvenir designed for the Society of Occupational Medicine, the British national organization for physicians in the field.

already surpassing the United Kingdom as centers of medical knowledge and research. American physicians studied pirated British textbooks (reprinted without royalties being paid to the authors or original publishers) and read British journals together with new and often transient medical journals that tended to be local and rather parochial. For the most part, American (and Canadian) medicine was a reflection of the British model of practice and something of a backwater, with a few conspicuous exceptions driven by ostentatious personalities such as

Benjamin Rush. However, there were differences that became more significant with time. Early American physicians certainly had patients who worked in mines and the early factories, just as did physicians in Britain, the mother country, but in the New World there was less of a social stigma attached to treating the working class. Also, because the United States and Canada had a perpetual labor shortage in the early years while the mother country usually faced a labor surplus (and dealt with it by emigration), the relationship between health, earning potential, and economic productivity was more obvious. General levels of health, to the extent that they are reflected in historical indicators, were also higher in North America than in the United Kingdom or continental Europe.

Occupational medicine may have been one of the earliest fields of medicine in the young United States to achieve sophistication. Benjamin McCready, a pioneer in public health and occupational medicine, published a famous paper on occupational diseases in the United States in 1837, just after Thackrah's landmark book appeared in Britain. The first American publication of a case study of phossy jaw (a degenerative disease of the bone in the mandible caused by white phosphorus exposure in match production) was in 1851, not long after the first papers on the disease appeared in Germany. The greatest physician of his era, William Osler, knew about Ramazzini (as did earlier well-trained physicians in the United States) and made his own contribution on the pathophysiology of coal workers' pneumoconiosis, the only experimental work he ever published.

By the middle of the nineteenth century, the rapidly growing cities of North America were following the industrial cities of England, Scotland, France, and Germany into what increasingly appeared to be a desperate social and economic trap: inadequate infrastructure supported an increasingly large population of first-generation urban dwellers pursuing individual livelihoods in a laissez-faire economy subject to wide swings in boom and bust cycles and periodic recessions, from which families never quite recovered. The result was a characteristic set of urban problems repeated in city after city and in industrial towns: a permanently impoverished lower class, a large criminal element, prostitution, child labor (which the family relied

upon for income), severe crowding in poor housing (with the attendant risk of tuberculosis), inadequate sanitation (leading to frequent outbreaks of cholera, other diarrheal diseases, and typhoid), alcoholism (and a rising opium problem), consumer fraud (particularly adulterated food), and chemical pollution, which at the time was recognized but not fully appreciated as a health risk. In continental Europe, there were revolutionary outbreaks in many major cities in the middle of the nineteenth century. These were met by violent repression but also by reforms intended to remove the causes of dissatisfaction. The United Kingdom, Germany, and, especially, the Austro-Hungarian Empire, embarked on ambitious social and health reforms designed in part to ameliorate conditions and in part to calm things down in order to achieve political and social stability.

Public health increasingly became a political priority in the United Kingdom during this time, culminating in the so-called Sanitary Revolution. The landmark Public Health Act of Parliament, in 1848, and later amendments and acts that strengthened it, established the management of public health as a state function in the United Kingdom. These foundational laws were followed by a series of legislative acts that introduced factory inspection, occupational health, pollution control, and control of child labor. Inadequate as they might have been by today's standards, they represented a great advance at the time and spread essential public health interventions throughout the Commonwealth. Although the achievements of the German and French public health movements were also known, the British public health movement was most influential and most copied in the United States and Canada.

The American commitment to public health developed more slowly, in part because there was less risk of revolution and in part because of the perception that the United States was better off than Europe, which had been true earlier in its history. In order to push for reforms, public health reformers such as McCready had to document that conditions in the United States were not so much better than in Europe and needed to be addressed on American soil. Public health advances in the United States were also delayed by the need to legislate state by state, but eventually public health took similar

shape in states such as Massachusetts. Occupational health, however, did not follow suit. Occupational health remained backward in the United States compared to Britain until the turn of the twentieth century. Later, Hamilton and others also drew on German models, which were more advanced, especially in research and prevention.

Sir Thomas Morrison Legge (1863–1932) was a highly influential British occupational physician who laid the foundation in the United Kingdom for modern occupational health regulation. His influence was strongly felt in North America as well. Dr. Legge had an unusual career path for his time. After Oxford and his medical education (St. Bartholomew's, London) he trained in public health (at Cambridge) and devoted his early career to preventive medicine and public health, particularly food safety. On July 26, 1898, he was appointed the first Medical Inspector of Factories, shaping that position into a leadership role in occupational health in Britain. For the next twenty-eight years, he built the Factory Inspectorate into the world's first modern occupational health and safety agency (the successor agency in the United Kingdom is the Health and Safety Executive). He was renowned for his erudition in the arts and history as well as occupational health and was considered an outstanding lecturer. He conducted original research on lead toxicity, anthrax (a hazard in industries using horsehair at the time), phosphorus, solvents, and numerous other hazards. After he retired because of ill health, he continued until his death to work as an adviser on occupational health to the trade union movement. He was greatly respected during his lifetime and was knighted for his work. He is primarily remembered for having pioneered effective occupational health and safety regulation. He was also famous at the time for “Legge's Aphorisms,” a series of pithy statements that guided occupational health practice (see Table 1.1).

### Railroad Surgery and Medicine

Occupational medicine practice in the United States arose from several roots, each representing a mode of medical practice in vari-

**Table 1.1.** The Aphorisms of Sir Thomas Legge (1934)

- 
1. “Unless and until the employer has done everything—and everything means a good deal—the workman can do next to nothing to protect himself, although he is naturally willing enough to do his share.”
  2. “If you can bring some influence to bear external to the workman (that is, one over which he has no control) you will be successful; and if you cannot, or do not, you will never be wholly successful.”
  3. “Practically all industrial lead poisoning is due to the inhalation of dust and fumes; and if you stop their inhalation you will stop the poisoning.”
  4. “All workmen should be told something of the danger of the material with which they come into contact and not be left to find it out for themselves—sometimes at the cost of their lives.”
- 

ous industries or in communities of workers. One of the most important was railroad surgery and medicine, which also originated in the middle years of the nineteenth century.

Railroad surgeons (the field was originally dominated by trauma care) emerged early in the nineteenth century as a distinct specialization in medicine (although before medical specialties were recognized). These surgeons and physicians were well organized, with medical societies (two major competing organizations and several minor ones), journals, and their own annual meetings. Railroad medicine and surgery became quite advanced on many levels between 1849 and the early twentieth century, becoming a leader in health-care. Railroads introduced an affordable prepaid medical care plan that was the original model for the Kaiser health plan and later for health maintenance organizations. Railroad surgeons provided superior care to injured workers, advised the railroads on sanitation, and introduced innovations such as emergency packs and mobile clinic cars. Many of these railroad physicians and surgeons were women at a time when women seldom pursued medical careers.

Railroad hospitals were well equipped and had worker representation on their boards. They left a legacy of access to healthcare in the

form of complete healthcare systems (such as Guthrie Health in eastern Pennsylvania), major medical center facilities in urban centers (in cities such as Chicago and Sacramento), smaller hospitals in towns throughout the country (such as Paducah, KY, and Clifton Forge, VA), and community-based care, which they supported through a retainer system that subsidized physicians who also provided general medical care in remote and marginal areas. Many railroad hospitals were teaching institutions. The Scott & White health system in Texas, for example, got its start as a railroad hospital and today continues the tradition of postgraduate training as part of the Texas A&M medical system.

Railroad medicine and surgery survived well into the twentieth century. However, despite the progressive features noted above, railroad medicine acquired a reputation of being biased against the interests of workers. The major reason for this appears to have been that their employers used railroad physicians as defense experts and in evaluating disability in passenger-initiated legal actions and workers' claims for compensation. Their opinions were considered to have been bought. Another reason was that the railroads charged their workers a small capitation fee for medical services. In an era when people rarely went to the physician unless they were acutely ill, and so did not expect to use the physician's services, this mandatory payment was resented.

### **The Progressive Era and Beyond**

In the early 1900s, two giants in the field emerged in the United States: the Progressive pioneer Alice Hamilton and Harry Mock, the model of an enlightened corporate medical director. The early 1900s also brought a major reform, workers' compensation, based on a German model.

By the last half of the nineteenth century, the havoc caused by the Industrial Revolution was recognized, and governments in Europe were facing a social and political crisis. Otto von Bismarck, the German chancellor, created in 1884 a system of social insurance that

included a seemingly generous pension system (less generous than it might seem because life expectancy after retirement was short then) and a system for income replacement and medical care for workers injured on the job. The United Kingdom and the Austro-Hungarian Empire created similar systems, which came to be called “workmen’s compensation,” although the United Kingdom later dropped its system when the National Health Service was introduced. Beginning with Wisconsin in 1911 and, in Canada, Ontario in 1913, workers’ compensation legislation was adopted throughout North America in the early twentieth century.

Also in 1911, a tragic fire at a sweatshop in New York, the Triangle Shirtwaist Company, killed 146 people, mostly young women. The doors had been barred by the operators in order to prevent unauthorized absence. Some of the victims jumped to their death in a desperate effort to escape the fire, a detail that particularly horrified New Yorkers. The incident drew national attention, however briefly, to working conditions in the United States.

Alice Hamilton (1869–1970) was the most influential occupational physician in North America in her day and had the same shaping influence in the United States as Legge in the United Kingdom (Figure 1.5). She is considered to be the founder of American regulatory toxicology and occupational (industrial) hygiene. On several occasions she served as a special investigator for the U.S. federal government, documenting occupational health problems. She is remembered as a woman of strong principle, a superb teacher, and a technically expert clinical toxicologist with a particular interest in lead. Hamilton was the first woman appointed to the faculty at Harvard and was well known in her day as a peace activist and reformer as well as a scientist. She was a disciple of Jane Addams and began her career in Chicago at the Hull House institution, one of the most influential reform movements in the mainstream of American life in the late nineteenth century. Hamilton was the first in a long line of women activists in occupational health who achieved much and made a difference in what was then very much a man’s world of factories and laboratories. Other prominent women in the history of



**Figure 1.5.** Alice Hamilton, the model for the progressive occupational physician and activist. (Image courtesy of National Library of Medicine.)

the field are Harriet Hardy (who collaborated with Hamilton) and Anna Baetjer (an extraordinary occupational hygienist). Hamilton's autobiography, *Exploring the Dangerous Trades*, inspired a generation of activists and continues to inspire progressive occupational physicians.

Harry E. Mock (1880–1959) was a distinguished teacher, clinician, and thinker in occupational medicine who established the model for the enlightened corporate medical director in the United States. Among other achievements, he founded one of the first clinics in occupational medicine to operate in an academic medical center, at Rush Medical College in Chicago. Mock drew on his nine years of experience as chief surgeon at Sears, Roebuck and Co. to produce an authoritative and prescient book, *Industrial Medicine and Surgery*, in 1919. Unusual for its time, the work develops a general theory of occupational medicine without reference to particular industries. Instead, it emphasizes periodic health surveillance, detection and

treatment of tuberculosis and sexually transmitted diseases among employees, injury prevention, rehabilitation of the disabled, and the particular needs of women and immigrants with respect to both occupational health protection and lifestyle, what would now be considered health promotion. In 1916 Mock was instrumental in founding the Association of Industrial Physicians and Surgeons, which ultimately became the American College of Occupational and Environmental Medicine.

The Depression, of course, was devastating to industry worldwide, and occupational physicians lost a great deal of their influence. Together with the workers under their care, many of them lost their jobs. Occupational health dropped off the national agenda. Regulation was considered to be an impediment to economic recovery. Workers accepted risks because they were grateful just to have jobs, and employers had no incentive to improve conditions. The Second World War and the earlier mobilization undertaken by President Roosevelt in anticipation of combat ended this situation, at least in industries with military significance.

### **Mid-Twentieth Century**

Occupational (then “industrial”) medicine peaked in the World War II years. During the War years, physicians with knowledge of technology, industrial production, toxicology, and work physiology were considered to be essential to the war effort. After the war ended, most leaders in occupational medicine at the time assumed that the field would maintain its stature. Until about 1950, the field was highly regarded and was well on its way to being recognized in American medicine, and accordingly its leaders made plans for a specialty board, to be named “occupational medicine” rather than “industrial medicine.”

However, occupational medicine suffered a devastating setback around 1950. Its precipitous decline in prestige from the war years is inextricably linked to battles within the American Medical Association that related to McCarthyism, the ill-fated initiative led by President Truman to create national health insurance in the United

States, and fear of “socialized medicine.” Occupational medicine got caught in a backlash against organized medical practice and rising interest in alternate forms of providing medical care.

In the 1950s there was a ferocious battle, mostly fought within the American Medical Association, between general practitioners (before family medicine was organized) on the one side, who believed that the only ethical form of medical practice was individual fee-for-service, and the smaller and (then) much weaker faction espousing medical group practices, prepaid health services, and occupational medicine, all of which were considered the “corporatist practice of medicine” (sometimes “corporate”), by which was meant the organization of medical care into a rational “corporate” system of managed healthcare, not medicine practiced by corporations. The corporatist practice of medicine was considered to undermine the integrity of the physician-patient relationship, part of which was the confidence built and expressed through direct payment of fees for services rendered. Of course, it was also highly threatening to solo practitioners, who feared competition if plant- or facility-based physicians provided primary care for workers and their families, a model that had in fact been introduced by some large employers.

In the heat of the argument, occupational medicine, other organized and managed forms of healthcare delivery (such as group practices and prepaid healthcare), and the concept of national health insurance were all denounced as unethical. Their opponents saw in them the insidious face of socialized medicine. When it was over, occupational physicians felt themselves to be outcasts and were deeply discouraged. Organized occupational medicine was outflanked on the right by the AMA, which created its own duplicative council to represent the field, and on the left by the American Public Health Association (APHA), the leaders of which wanted occupational medicine to be integrated into public health services and wanted a worker-oriented occupational health system to be part of a national healthcare system; these leaders often felt that occupational medicine practitioners were not radical enough politically. (The

AMA and the APHA were also at odds on other issues, but the big fight between them was over national health insurance.) Specialty certification in occupational medicine was initially put on hold, due to the AMA's opposition, but was later achieved, after various compromises, through an uneasy merger of convenience with the fields of preventive medicine (which was more closely identified with organized public health than it is today) and, more comfortably, aerospace medicine.

In the long run, of course, advocates for organized and managed healthcare won control. Managed care organizations got bigger, became financially dominant, and took over as the driving force in American medicine. HMOs, medical groups, prepaid practice plans, and the whole range of managed care organizations (MCOs) and systems eventually were accepted by physicians because they had to be. The MCOs forced themselves into position as the dominant mode of healthcare delivery in the United States. Once they were in control, organized medicine could do nothing about it. In the event, the HMOs and other MCOs bypassed medical authority, commodified healthcare, and turned physicians into employees.

However, the ultimate if begrudging acceptance of managed care and the need for other models of health insurance did not extend to acceptance of occupational medicine as a form of organized and managed healthcare. Without clout or prestige, occupational medicine stayed out in the cold for many long years. Practitioners in the field continued to be treated as deviant or suspect by the rest of medicine, even if the reasons for the prejudice were largely forgotten. That is why both the specialty and the field of practice generally developed in relative isolation for so many years, insecure, with strained but correct relations with the mainstream of organized medicine.

Other countries followed very different paths and pursued models that put occupational medicine at the center of healthcare and worker protection. One of the most important models was the system of organization of occupational healthcare in Sweden. Sven P. M. Forssman (1911–1990) was the architect of the occupational

healthcare system in that country, which was considered by many to be a model for the world. A physician and biochemist by training, Forssman embarked on a career in occupational medicine very early and by the age of thirty-two (in 1943) was a full Professor of Occupational Medicine. After a long and productive career, he began working with employers, labor unions, and the government in Sweden to improve occupational health services. The result was a collective agreement in 1966 that created a radically new network of clinics, referral centers, training programs, and information centers covering the entire country, accessible to all employers and workers, and based mostly at universities in order to ensure quality of care and research into occupational health problems.

The system functioned effectively for twenty-five years, during which Sweden became a world leader in the field, as it had in many social issues. However, in 1990, the highly successful Swedish economic model stumbled and the country experienced a deep recession, although it was ultimately short-lived. The occupational health system came to be perceived (many would say misperceived) as overbuilt—overly elaborate and too expensive for Swedish industry in a time of recession. It was essentially dismantled in the political reforms that took place in Sweden in 1992 following the election of the Conservative government after years of rule by the Social Democrats. The employers' confederation unilaterally withdrew from the pact, knowing that the government of the day would not oppose the move. The result was rapid disintegration and dismantling of the world's premiere example of a voluntary occupational health protection and management system. Sweden then found that it had relied so heavily on the pact, on the assumption that the model would continue forever, that it lacked the government infrastructure, policies, and regulatory regime to manage occupational health issues. In 2007 Sweden closed its National Institute for Working Life, which had subsumed the National Institute of Occupational Health, of which Forssman was the first director.

International leadership then largely passed to Finland, which in the meantime had maintained and grown the Finnish Institute for

Occupational Health into a position of world leadership, under the visionary direction of Jorma Rantanen.

### **Approaching the Present Day**

During its period of contraction, occupational medicine in the United States was kept vigorous mostly by corporate medical departments and research departments, the dedication of a small number of academic occupational physicians, and a few university programs.

Chief among its advocates was Robert Kehoe (1893–1992). At the University of Cincinnati, he served as professor and director of the Kettering Laboratory, which is still a major institution in academic occupational medicine and his legacy. Kehoe was a renowned toxicologist and taught scientific rigor, but much of his work has been discredited and his reputation has suffered accordingly. On the positive side, he was a leader in academic occupational medicine at a time when it was under grave stress and might not have survived without his strong personality. However, he is remembered today largely as an apologist for the lead industry. Although he acknowledged the risk to children from lead paint, Kehoe adamantly denied that airborne lead presented an environmental hazard. He served as chief medical consultant to the Ethyl Corporation, which produced the tetraethyl lead that was added to gasoline as an antiknock compound, from 1925 to 1952 and continued his association informally for the rest of his long life. In both his position as corporate consultant and in his scientific work, he propounded the view that environmental sources of lead presented a minimal hazard to human health. In fairness, it should be noted that he was personally convinced that lead was an essential trace element with a normally high background concentration in the general population and was therefore unlikely to be highly toxic. However, he clung to this belief long after the weight of the scientific evidence had turned the other way. Because there is no indication that he did not believe what he said, he may be regarded as a tragic figure, a scientific giant with clay feet.

By the late twentieth century, British occupational medicine was less important in shaping occupational medicine in North America, but retained important influence, especially in Canada. Developments and policies in the United Kingdom helped to validate and support similar developments in the United States, and academic occupational medicine continued to influence and inspire the considerably diminished academic infrastructure of occupational medicine in the United States. During the mid- to late twentieth century, the United Kingdom produced key leaders in occupational medicine whose legacy is still felt worldwide, especially through the many students they taught.

Dr. Donald Hunter (1898–1978) was a one-man institution in British occupational medicine: director of the Medical Research Council Department of Research in Industrial Medicine, founding editor of the *British Journal of Industrial Medicine* (renamed *Occupational and Environmental Medicine*), consulting physician at the London Hospital, and ubiquitous lecturer and author of *Diseases of Occupations* (still in publication, now prepared by a team of authors). A curmudgeon known for his sharp tongue, Dr. Hunter established an incomparable standard of scholarship (particularly in the history of the field) but also wrote extensively for the public, in order to promote understanding of occupational health and its importance in society. His successor, Richard Schilling, was a beloved figure in British occupational medicine, known for his deep empathy, competence in practical matters, and excellence as a teacher and mentor.

The Occupational Health and Safety Act of 1970 changed the direction of occupational medicine in the United States. The act created the Occupational Health and Safety Administration (OSHA) within the Department of Labor and gave it a legal mandate and limited power to regulate occupational health. This lent new urgency and visibility to occupational health, although the backlash against OSHA and the intrinsic weaknesses of the agency soon would compromise much of its effectiveness. The act also created a new and different entity, the National Institute of Occupational Safety and Health (NIOSH). Formed initially from preexisting laboratories and

bureaus of the U.S. Public Health Service and the Bureau of Mines, NIOSH developed a role as the primary federal sponsor of research and training. Its training grants and support for multi-profession “Education and Research Centers” created and sustained a new foundation and infrastructure for the principal fields of occupational health, including occupational medicine. For the first time, occupational medicine training programs, as residencies and later as fellowships, could be assured of sustained support apart from corporate contributions. NIOSH, through its support of training programs, has now brought two generations of physicians into occupational medicine. The earliest of them largely saw themselves as working in the tradition of public health and worker protection, rather than under corporate auspices. They found a role model in a middle-aged pulmonary physician from New York.

Irving J. Selikoff (1915–1992) was the single most influential figure in American occupational medicine after the Second World War. He was a powerhouse of energy and activity. Before becoming interested in occupational lung disease, Selikoff had a distinguished career as a consultant in pulmonary medicine in New York, where he conducted the clinical trials that established isoniazid as effective treatment for tuberculosis. Relatively late in his career (he was about fifty), starting in the 1970s he embarked on a long series of imaginative and technically difficult studies that conclusively documented the risks of exposure to asbestos and the interaction between cigarette smoking and asbestos exposure in causing lung cancer. Having established asbestos as a major hazard and using it as a model of an occupational hazard, he conducted broader studies in occupational cancer and later in other occupational and environmental health problems, such as the polybrominated biphenyls (flame retardants, no longer used), in an ever-widening scope of interest. He founded the occupational medicine program at Mt. Sinai School of Medicine and trained a generation of activist physicians there and by example encouraged progressive occupational physicians across the country. Many of his students and disciples have become leaders in their own right. He also founded or cofounded an astonishing number of national and international

organizations (at least five that remain active), headed many more, founded two of the leading medical journals in the field, edited eleven books, and wrote 350 papers, several of which are considered classics in occupational medicine. He could be a very polarizing figure and provoked strong reactions in people. Sometimes these reactions were petty. Shortly after his death, for example, there were efforts to discredit him by calling into question the legitimacy of his medical degree, which he had obtained in Scotland because there were quotas on the admissions of Jews to American medical schools before the Second World War. On the other hand, Selikoff was known for showing tolerance and respect for those who genuinely disagreed with him and who could argue their case scientifically.

No leader in occupational medicine has emerged since Selikoff with the same charisma, gravitas, and influence. Although science has advanced and there have been many discoveries, none have had the impact of asbestos, both as an individual hazard and as a near-perfect model for understanding occupational disease. Instead, the course of occupational medicine has been determined, buffeted even, by events largely out of its control, mostly involving changing business management trends.

Around 1980, new management philosophies and trends became popular that emphasized reducing the size of companies (“downsizing”) and removing layers of management (“delayering”) in the interest of efficiency. Activities that did not directly relate to the core business of a company were subject to “outsourcing,” to be provided by contractors.

The result of these trends was to drastically reduce the commitment of major corporations to internal employee health programs. Many corporate medical departments disappeared entirely and community-based occupational health services sprang up to provide services at the plant level on contract. Because the supply of specialist-trained physicians has never been sufficient in occupational medicine, the demand was largely met by physicians moving into the field from family practice, emergency medicine, or other specialties. For a time, the more specialized and technical aspects of occupational medicine, such as tox-

icology, were often dropped, reduced in complexity, or, rarely, outsourced to highly specialized consultants. In-plant services were often turned over to occupational health nurses and health promotion, and employee assistance programs came to be run by nonmedical managers. An earlier trend toward consolidating environmental health issues and consumer protection with corporate occupational health was reversed, and employers let separate contracts for various services, which were then less well coordinated.

Surprisingly, indicators such as membership in ACOEM show that during this period the number of physicians active in occupational medicine actually expanded. There appear to have been more positions in occupational medicine created in the community than were lost by the closing of corporate medical departments and plant facilities. The increased availability of occupational health services in the community may also have created a new demand for services by mid-sized enterprises that could not afford medical departments of their own.

There were negative outcomes from the outsourcing trend. The medical services that remained on the table tended to be acute care and workers' compensation management, so the field may be said to have "de-specialized"; it was made simpler, more routine, and less in need of deep expertise. Downsizing and delayering made prevention-oriented services much more difficult for occupational physicians to provide and reinforced the trend toward acute care. However, the greatest disruption came as an unintended consequence of delayering. In time, many of the mid-level managers and human resources officers who had previously managed contracts for outsourced occupational health services eventually disappeared themselves, leaving a vacuum in the corporate management of health affairs. The new breed of MBA-trained managers who entered business during this period had little or no instruction in business school on health issues and no experience in managing occupational health services.

In recent years, the market for occupational health services has firmed and is supported by several trends. Some employers, perceiving that their management of health affairs was weak, have brought

back a small number of OEM physicians. Large-scale retirement in some industry sectors and in the U.S. federal government has created new demand. New trends in business, such as the trend to open primary care clinics on-site, have created new opportunities. Although not a harbinger of massive growth, current trends suggest an expansion in coming years that cannot be met by OEM specialists because of the bottleneck in training capacity.

## **ENVIRONMENTAL MEDICINE**

There is much work left to be done in integrating environmental medicine fully within OEM. Environmental medicine today continues to have a fuzzy image, a confused identity, and an indistinct scope of practice. The American College of Occupational and Environmental Medicine has established competencies for environmental medicine within OEM (see Chapter 12), but they are not well known, even within OEM. The American Board of Medical Specialties (the authoritative body in such matters) has never recognized it as a distinct specialty, although for many years it has been (appropriately) consolidated within the specialty of occupational medicine by the American Board of Preventive Medicine. However, ABPM does not use the term “environmental” in the name of the specialty, which reduces recognition of these competencies in the field.

Given this vacuum of recognition, various practitioners and schools of thought have appropriated the term “environmental medicine” as their own and have further confused its identity. The confusion is reinforced by alternative medicine advocates and by misinformation, especially on the Web.

Historically, occupational physicians tended to be called in on problems involving air quality and hazardous waste, because the chemical regime tended to be similar to that in the workplace. One example was the Donora air pollution episode of 1948, which was investigated by occupational physicians and hygienists from the occupational health branch of the U.S. Public Health Service, as it was called at the time. General preventive medicine and (medical)

public health specialists, on the other hand, tended to be called in on issues involving water and food safety because their core expertise was public health protection and more often involved infectious disease. This division of interests has changed with consolidation of responsibilities among consultants and medical directors.

### **Practicing Environmental Medicine**

OEM physicians are expected to know about environmental influences on health and, increasingly, to have management skills that apply to environmental hazards and consumer protection as well as occupational health protection. The American Board of Preventive Medicine includes core content in environmental medicine in the board certification examination for occupational medicine.

Clients and employers need evidence-based, knowledgeable expertise in dealing with medical aspects of environmental health issues. It took them a long time to discover that this is within the scope of OEM, but at least in major corporations and high-risk enterprises, which face potentially serious liability issues if they do not manage their environmental responsibilities well, the versatility of OEM physicians is recognized.

Practitioners, medical directors, and consultants must deal with an increasing range of activities that fall into the domain of environmental health: managing environmental hazards (traditionally, exposures that occur “outside the fence” or “beyond the plant boundary”), regulatory compliance, consumer safety and protection and the concomitant problem of product stewardship throughout its life cycle, legacy liabilities (as a result of past operations or the merger with or acquisition of companies with significant liabilities), litigation, globalization and environmental issues in the sector, responsibility for foreign operations, emerging infections in the workplace (such as SARS and pandemic influenza), and public health protection in the workplace, for example in food services. For this level of service to institutions and on the level of populations, risk communication is a critical skill (see Chapter 7).

As obvious as it is in the corporate and government sectors that occupational and environmental medicine is all of a piece, the practice of environmental medicine in the private sector and in clinical practice is not as easily defined. Allergists practice a form of environmental medicine, and atopy drives many more cases of “environmental” disease than occupational disease. A working knowledge of allergies and immunology is therefore essential to the practice of environmental medicine. Medical toxicologists are concerned with environmental exposures, but this is not the primary focus of the specialty, which is driven in practice by emergency care and drug overdoses.

Environmental medicine is a natural extension of the clinical scope of practice of occupational medicine, but in private practice the business side is very different. Environmental medical services are not viable unless subsidized or combined with the practice of other modes of medicine. Occupational medicine clinics run the gamut of relevant disorders, and patients come to a clinic through contracts or employer agreements, referrals, labor organizations, government agencies, independent medical examination work, and self-referral. These mechanisms for patient referral have no counterpart in environmental medicine. Payers, such as insurance companies, are usually very resistant to approving referrals and resist payment for environmental medicine services. This is due to several understandable reasons: lack of familiarity, suspicion of irrelevancy, confusion with alternative medical practice in “clinical ecology” and junk science, lack of tangible benefit (diagnosis is not as impressive as a procedure), rejection of cases that could be covered by workers’ compensation (if there is any question of a workplace connection or if the physician self-identifies as exclusively practicing occupational medicine), and disallowing services for purposes of litigation.

The flow of patients is also limited. In two general university OEM consultation clinics—institutional members of the Association of Occupational and Environmental Clinics—approximately 15 percent of cases were “environmental” rather than occupational in that the chief complaint did not relate to an exposure in the workplace. Most of these cases were self-referred and involved sick building syndrome or indoor air quality, the interpretation of trace element analyses not

related to workplace exposures, mold (and concern over supposed mycotoxin effects, which always had another explanation), and pesticides. This is not enough to support a viable practice outside of an academic setting, where costs can be cross-subsidized. OEM physicians in the mainstream of medicine who wish to develop a practice predominantly in environmental medicine will therefore have difficulty supporting themselves through patient care and evaluation alone. (Physicians who practice alternative medicine, in the form of “clinical ecology” or other unproven schools of thought, are considered later in this discussion.)

Involvement in medicolegal work is an opportunity for the provision of medical-expert, although not strictly medical, services. There is a heavy demand for evidence-based litigation support and testimony on topics such as pesticide exposure, mold, groundwater contamination, and many other environmental hazards overlapping with consumer product safety. This is not easy and is certainly not for the casual or untrained practitioner. The physician who gets involved will quickly be tested in terms of expertise, rigor of analysis, fund of knowledge, and partisanship (plaintiff or defense orientation). On the other hand, ethical and evidence-based medical-expert practice is a way of applying medical knowledge in nontraditional ways and contributing to the resolution of disputes involving torts. (See Chapter 23.)

Complicating the practice of environmental medicine is the reality that the name tends to attract patients with a preconceived idea of what the field is. Rather than an evidence-based medical and public health approach, some patients who self-refer are looking for a spiritual or integrative experience, as captured by the common term “holistic.” These same patients are often consciously looking for practitioners who challenge the validity of mainstream medicine, and those patients will be disappointed by an approach that appears to conform to scientific medical practice and rigor. It is often helpful for the practitioner to have a statement on his or her Web site and to have the staff state clearly at the time of the appointment that the practitioner is not an adherent of alternative medicine; this will avoid disappointment and allow patients who are looking for a sympathetic practitioner to shop elsewhere. Other patients are looking for a plausible explanation for

their problems and do not much care what the theory behind the diagnosis may be. These individuals may be equally happy with a diagnosis of “multiple chemical sensitivity” as a diagnosis of garden-variety allergic rhinitis.

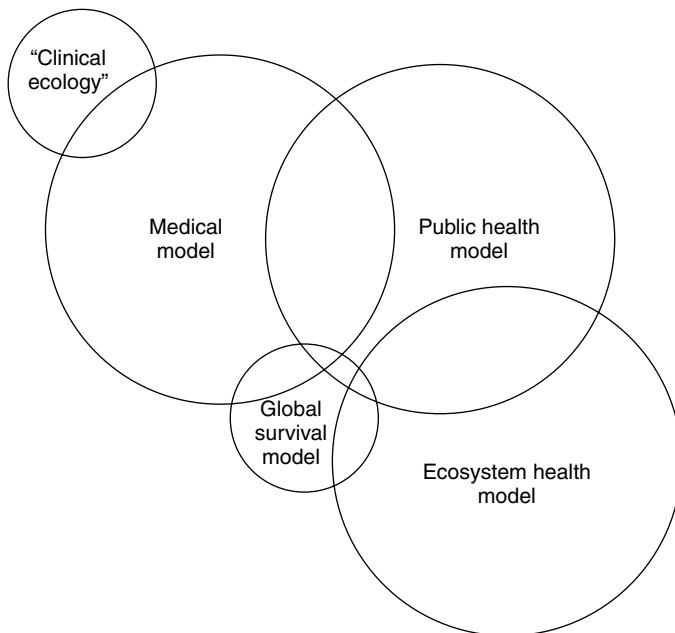
## **Concepts of Environmental Medicine**

“Environmental medicine” means different things to different people. At least five distinct schools of thought can be identified, of which the medical model and the public health model form the core of the environmental side of OEM. Each is associated with a set of champions and advocates, organizations and priority or defining issues. Each attempts to be comprehensive in scope within its own paradigm. The five are schematically represented in Figure 1.6, a Venn diagram that shows their degree of overlap. The approach in this book generally combines the medical model and the public health model.

### ***The Medical Model***

Mainstream physicians who are concerned with environmental diseases as clinical entities and the diagnosis of individual cases of disease generally come from a medical tradition that grows out of occupational medicine and internal medicine. The emphasis is on environmental exposures as specific or contributory determinants of defined clinical disorders, subject to diagnosis and management using mostly conventional clinical methods and approaches. Toxicology and ergonomics tend to dominate in the literature, and the methods of clinical investigation are better suited to its issues. Clinical cases are managed one at a time; their numbers are small; and they do not necessarily arise from a defined population.

The major institutional representatives of this school of thought in the United States have been the Association of Occupational and Environmental Clinics and the American College of Occupational and Environmental Medicine. NIOSH and ATSDR have been leading



**Figure 1.6.**Venn diagram of the various schools of thought in environmental medicine. Mainstream medicine rejects “clinical ecology.” (Reproduced from “Of Blindmen, Elephants, and Environmental Medicine” by Tee L. Guidotti. *New Solutions: A Journal of Environmental and Occupational Health Policy*, Volume 6, No. 4, pp. 25–30 [Summer 1996]. Reprinted with permission.)

sources of support for this school of thought. The institutional base tends to be in medical schools and community-based clinics.

### ***The Public Health Model***

A relatively small number of OEM physicians work in the professional public health sector or in population-based research. These physicians do see environmental medicine as the medical approach to traditional environmental health concerns, including disorders arising from problems of water quality, air quality, and sanitation. This school of thought tends to be “hazard-driven” in the sense that the line of reasoning starts with the hazard and its possible effects or refers back to a particular hazard as a cause of disease in the population.

Although this school of thought shares with the medical model an emphasis on environmental hazards as the cause of disease, it relates exposure to disease patterns in the population rather than the individual and is largely concerned with the recognition and control of environmental exposure to prevent disease. Epidemiology is the principal discipline for physicians in this school of thought. Standards setting is a major preoccupation, as is monitoring and surveillance to assess the adequacy of regulatory standards. The major institutional “homes” for this school of thought in the United States have been schools of public health and public health-oriented organizations such as the American Public Health Association. Aside from textbooks of public health and environmental health, the best example of this school of thought in the medical literature is William Rom’s *Environmental and Occupational Medicine*, which is the most exposure oriented of the mainstream texts in occupational medicine.

### ***Ecosystem and Human Health***

This school of thought is explicitly adhered to by a very small number of physicians but a very large number of environmental activists and many biologists. Basically, it emphasizes the implications of ecosystem change for human health and society, on the very plausible grounds that all human life depends on an intact and biologically productive environment. Great emphasis is placed on sustainable development, the health implications of global change, and ecosystem integrity. Overpopulation is a consistent factor in the equation and the major concern of a subset of this “school.” Sometimes the link to human health is difficult to make or is indirect, and it is usually presented in the form of broad-brush scenarios for the future. Because there is no one biomedical or population-based discipline that applies (other than demography), research in this area is sparse, difficult, eclectic, and draws heavily on secondary analysis of findings from the environmental sciences generally and from demography.

This remains a less influential school of thought among physicians because their role in it is less clearly defined and because there is no

clear medical dimension to the problem. Environmentalists and the public, on the other hand, tend to attribute neglect of the issue by physicians to apathy, lack of concern, or professional conservatism. The guiding force in this school of thought is the International Association for Ecology and Health and its journal, *EcoHealth*.

### ***Ecosystem Damage as Global Violence***

This school of thought is really an ideology that grew out of the medical response to issues of nuclear war and the highly successful campaign to mobilize physicians as an effective force for world peace, International Physicians for the Prevention of Nuclear War. The radical critique that informed much of the opposition to the nuclear and defense establishment is extended in this analysis to environmental issues. The simplifying assumption is that ecosystem damage, like warfare, represents violence to the planet and all human societies and species. The assumption is attractive but is often inadequate to carry the analysis of issues much beyond initial awareness. It is parallel to and consistent with the line of thought about warfare as an ecological issue that one encounters within the World Health Organization. The leading representative of this point of view in the United States is Physicians for Social Responsibility, and in Canada, Physicians for Global Survival; the Union of Concerned Scientists is also highly influential. The critical text is *Critical Condition: Human Health and the Environment*, edited by Eric Chivian. This model of environmental medicine is passionately held by a small but important group of physician activists.

### ***Alternative Medicine***

Some practitioners who self-identify by the term “environmental medicine” belong to a school of thought that is not accepted by the mainstream of biomedical science but that became influential on the public and among many healthcare professionals in the 1980s. This school of thought has gone by different names but was known for many years as “clinical ecology.” It shares with the “medical model”

school an interest motivated by individual cases but bases itself on very different and constantly shifting concepts of etiology. It may be seen by some as an extreme or distorted view of the medical model, characterized by inadequate clinical research methods, exceptionally fluid case definitions, questionable logic, and untestable hypotheses, all of which are generally considered characteristics of “junk science,” not mainstream science.

Advocates for this point of view have attempted to appropriate the name “environmental medicine” by creating a free-standing certification board: The American Academy of Environmental Medicine. This group is driven primarily by the “American Environmental Health Foundation” and the teachings of Dr. William J. Rae in Dallas. In recent years the group seems to have lost adherents, and some of the energy has been diffused into advocacy for “toxic mold” as a disease entity and other questionable health theories. This school of thought has created an impediment to the development of environmental medicine because of the need to deal with the expectations of patients who believe that it, rather than the medical or public health models, represents the current state of medical knowledge.

Given this situation, an important function of mainstream OEM has been to serve as a line of defense against questionable practice. People tend to project their fears and beliefs about health onto their environment. As a consequence, OEM clinics and practitioners see more than their share of the worried and the obsessed and tend to be magnets for people who believe in multiple chemical sensitivity, dental amalgam disease, toxic mold, and extravagant claims for the effects of electromagnetic fields. The role of OEM physicians is frequently to explain patiently, to review the evidence, to say no, and sometimes to take abuse when the patient’s beliefs are not supported and his or her expectations for exotic care are not met.

### **Physicians as Advocates**

Physicians are also citizens and have opinions with respect to environmental issues. Some physicians, for example Anton

Chekhov (the playwright), have distinguished themselves as advocates for the environment. (Chekhov's special interests were conservation and reforestation.)

Studies of risk perception have repeatedly demonstrated that physicians are among the most credible sources of information on environmental issues, even those that are not driven by health concerns. There is nothing in conventional medical training that warrants this high respect, however. Knowing that, physicians have a responsibility to take great care in staying within the scope of their practice and expertise when giving opinions on issues of the environment. Physicians have the advantage of rigorous training and a deep knowledge of the human body.

Physicians in certain fields (allergy, medical toxicology, pulmonary medicine, infectious disease) are trained in and do develop specialized expertise on environmental health issues of direct relevance to their practice and patients. Primary care practitioners in the present day (not necessarily in the past) are certainly trained to be aware of environmental influences on health, but usually not to a degree of specialized expertise. Specialists who deal with diagnosis and treatment of the end result may be generally concerned about causes as a professional commitment (for example, almost all physicians are against cigarette smoking) but few have or seek expertise in causation or environmental management. Most physicians have no special expertise in this area. (For example, an oncologist must know diagnosis and treatment and usually knows the epidemiology of a cancer, but is usually not trained to analyze causation or to manage the risk of exposure to a carcinogen.) The three preventive medicine specialties—preventive medicine, occupational medicine (OEM), and aerospace medicine—together with medical toxicology do require such training. However, these four fields are each small and they are exceptions to the rule.

Organized medicine has been relatively slow to embrace environmental issues, but many societies are doing so now. Physicians are stakeholders and, as caregivers who see the consequences of bad health risks and the failure to control public health threats, have a

point of view in environmental issues. Many medical societies, such as the American Academy of Pediatrics, have committees or internal bodies devoted to environmental health that are equipped to deliberate on complex issues and to weigh scientific evidence. These organizations skillfully blend the specialized expertise of members who are well prepared on these issues with the interest and commitment of members in the community.

The conscientious physician has a responsibility to the profession and to society not to abuse the confidence and trust that is placed, or misplaced, in physicians. The guiding principle is always to be aware when one is speaking as a professional and when one is speaking as an individual or advocate, and to keep these roles clear, separated, and unambiguous when speaking in public. One way to achieve this is for medical and scientific organizations to stick to evidence-based issues and for medical policy or advocacy organizations to handle the opinion and advocacy work.

There are many environmental advocacy organizations specifically or predominantly for physicians worldwide, but not in the United States. Physicians for Social Responsibility in the United States and Physicians for Global Survival in Canada are advocacy organizations that began with opposition to nuclear war and have a diverse agenda that includes environmental advocacy. The Canadian Association of Physicians for the Environment (CAPE) is a highly influential organization that began with the mission of bringing subject matter experts together with physician activists in order to create informed advocacy and to express opinions. CAPE and similar organizations in other countries belong to a confederation known as the International Society of Doctors for the Environment (ISDE). There is no counterpart organization in the United States.

Individual physicians can certainly inform themselves and can become advocates for the environment like any citizen. Their opinions matter as much as or more than those of non-physician citizens, simply because physicians tend to be opinion shapers in their communities. Individual physicians can be powerful advocates for environmental integrity.

There are several ways that physicians can make a contribution to environmental awareness and environmental protection on a local level. They include the following:

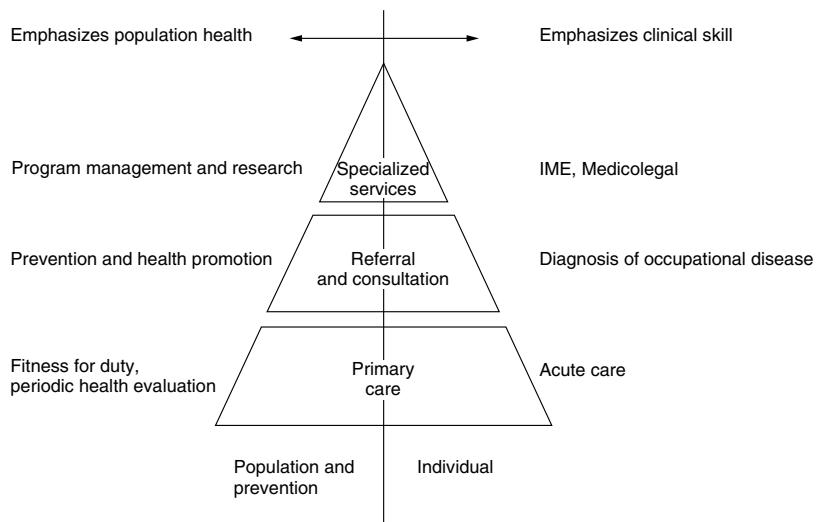
- Promote continuing medical education opportunities on relevant topics, such as waterborne disease, the health effects of air pollution, asthma triggers, hazards in the home, and the health effects anticipated to accompany climate change.
- Lead by example with green and eco-friendly offices and clinics.
- Support and promote the Green Hospitals and Health Care movement and its program of ecologically sustainable facilities.
- Support political action on health-related issues.
- Document environmental problems (with sufficient rigor to stand up to scrutiny).
- Advise environmental advocacy organizations on the findings and also the limits of medical knowledge, consistent with the physician's own expertise.

## **THE FUTURE OF OEM**

OEM practitioners increasingly work within a parallel healthcare system, with different capabilities and drivers than those that characterize the general healthcare system, sharing only providers, paid for by workers' compensation and employer-sponsored services (such as OSHA-mandated surveillance). During the 1980s and 1990s, the system became more clinically oriented than prevention oriented. The leaders of ACOEM have consciously sought to reconnect to public health and the principles of prevention, defining OEM physicians as "public health professionals for the employed population." However, because physicians are valued by society mostly for their clinical skills, there has been less opportunity for them to contribute to prevention, outside of research. The expansion of opportunities holds promise for a return of the primacy of prevention in the field.

OEM does not function on a single level of service, but instead constitutes a whole system that includes primary care, specialty care, and program management for groups. This is illustrated in Figure 1.7, which represents a pyramid of healthcare services. The bottom tier, supporting the others, is the primary care of occupational medicine; the next up represents referral and consultation services; and the top represents highly specialized services. Most OEM physicians function on one or, at most, two levels, on either or both sides. Also, few occupational health services are fully integrated. This means that almost every institution engaged in occupational health delivery sees it in a fragmentary way. A big challenge for the future is to present OEM as a whole.

OEM needs to take the lead in the work-related dimension of patient care, not necessarily by doing it all but by influencing physicians to do it right. Historically, most physicians have not done a very good job of assessing fitness for duty (writing uninformative “back to work” notes), assessing causation (without systematic analysis), evaluating disability (without justifying the conclusions), and managing sickness absence (such as writing “doctor’s notes” for patients after an



**Figure 1.7.** The pyramid of occupational health services.

illness that the physician never saw). Such services are highly unpopular but are essential to the patient, the employer, and the workers' compensation system. They are also often difficult for a physician who has a close relationship to the injured worker/patient and his or her family as that person's physician, with pressure to bend or stretch medical judgment on issues such as time away from work and level of impairment. These tasks are best done by a physician who is trained, is motivated to care about such seemingly unrewarding functions, knows the standard terminology and procedures of the system, and can be objective about the injured worker.

One area where OEM has certainly taken the lead is in issues of health and productivity. OEM physicians pioneered studies into the relationship between health and work performance. This important field will continue to grow despite perceptions by some that it subordinates the interests of the worker to production and the ends of management. This radical critique is only partially true. Certainly management is primarily concerned with the business of the enterprise, but that does not preclude a legitimate interest in the welfare of the worker. Most individual managers seek to do the right thing, and a deep knowledge of health and productivity makes the case that they need to do so for business purposes. In addition, protecting good health and fitness is not just a means of demand reduction to control medical costs and a tool to support productivity. As long as the policies are not coercive and privacy is protected, health and wellness programs are also a way to mobilize the power of the organization for the good of workers.

Today OEM in the United States has come out of a long down cycle and is reenergized, although the recent economic downturn has blunted its recovery. Demand for OEM trainees has been way up and is expected to expand, especially with projected U.S. federal government recruitment and the new trend, since the 1990s, of placing clinics in facilities and workplaces.

Also, the baggage heaped on occupational medicine by its detractors is on the decline. It has been over thirty years since any occupational physician in a position of influence, with the authority to be

listened to and quoted by the media, has said anything profoundly stupid that the public remembers. Two generations (in terms of social turnover) have passed, and neither the public in general nor colleagues in medicine remember why the field had an image problem in the past (or maybe even realize that it did).

Today the public, patients, colleagues, and other people OEM physicians deal with in management belong to a new generation, one that makes up its own mind and is not much interested in history. They know that medicine is not practiced in an economic and social vacuum. They have seen OEM practitioners try to do the right thing and get results. They are increasingly impressed. Tragically, a deep funk seems to be seeping through the rest of organized medicine, particularly in primary care, at the very time that OEM's future seems to be growing brighter.

OEM has been grappling with big, nasty issues for many years. Not surprisingly, the field has lots of hard-won experience, insight, and ideas about solutions and a long history of deep concern with ethics, balancing responsibilities (mostly among the worker/patient, the employer, and the government), and conflicts of interest. The rest of medicine is only beginning to come to grips with these issues, now that medicine is organized and managed, and physicians report to levels of supervision above them. OEM could be the vanguard of medical practice for the future: outcome oriented and evidence based, providing ambulatory care, remaining conscious of population impacts, valuing prevention, employing skillful case management, capably working with populations as well as individuals, receiving their education outside hospitals, and applying the knowledge of medicine broadly and beyond clinical practice.

What is next for occupational and environmental medicine? Workers' compensation reform, new technology with uncertain risks (such as nanomaterials), and new trends in healthcare are all favorable for the expansion and sustainability of the field. As of this writing (mid-2009), the future effects of the economic recession are not clear, but demand for OEM services is still increasing. Growth in trained specialists has been severely constrained by a choked training pipeline, which will make it impossible to supply the numbers the

field needs. The field will therefore continue to depend on talented, motivated physicians who enter in mid-career.

The future should bring a realignment between OEM and public health, which has been strained in recent years. Today OEM practice is informed by a simple idea: "Occupational and environmental physicians are public health professionals; occupational medicine for the employed population and environmental medicine for everyone." OEM physicians are indeed public health professionals and belong in close affiliation with public health and preventive medicine.

OEM has a past to be proud of. Although little recognized, it has been in the vanguard of the most important trend in medical care in the twentieth century: the organization of healthcare into managed and integrated systems. It learned early on of the power of large organizations to make health gains, to support and promote good health among communities and populations, and to support healthful behavior and good outcomes for individuals. Not many people were thinking that way in the 1940s and 1950s, but the leaders of occupational (and environmental) medicine were.

OEM physicians have become leaders in bringing high-quality medical care, thoughtful case management, and accountability in workers' compensation. A few years ago, those descriptors would not be found in the same sentence. In particular, the American College of Occupational and Environmental Medicine is at the forefront of advancing evidence-based medicine, especially through the ACOEM *Practice Guidelines*.

OEM has taken the initiative on issues of excellence in healthcare, health and productivity, and workforce protection, and has done so quite successfully. Good occupational medicine, including disability management and integrated care, can improve the lives and hopes of injured workers and their families. Health and productivity are not just about increasing productivity for the employer; they are also about improving the personal lives of workers and their families. Good environmental medicine identifies and helps solve problems and commits the physician to a better and more healthful world. OEM puts all these values into action.

## RESOURCES

- Gochfeld M. Chronological history of occupational medicine. *J Occup Environ Med.* 2005;47(2):96–114.
- McCunney RJ. *A Practical Approach to Occupational and Environmental Medicine.* Philadelphia: Lippincott Williams & Wilkins; 2003.
- Walsh DC. *Corporate Physicians: Between Medicine and Management.* New Haven, CT: Yale University Press; 1987. [Excellent analysis of the inherent tensions experienced by the subset of OEM physicians who work in corporate positions.]

## NOTEWORTHY READINGS

- Aldrich M. *Death Rode the Rails: American Railroad Accidents and Safety 1929–1965.* Baltimore, MD: The Johns Hopkins University Press; 2006.
- Bartrip PWJ. *The Home Office and the Dangerous Trades: Regulating Occupational Disease in Victorian Britain.* Amsterdam: Rodopi; 2002. Clio Medica 68.
- Draper E. *The Company Doctor: Risk, Responsibility and Professionalism.* New York: Russell Sage Foundation; 2003. [A disturbingly bad book, filled with misinformation and half-truths; mentioned here as a warning.]
- Gillespie R. *A History of Railway Surgery.* <http://railwaysurgery.org/>. Accessed January 2, 2008.
- Guidotti TL. Of blind men, elephants, and environmental medicine. *New Solutions.* Summer 1996;6(4):25–30. (Source of Figure 1.6.)
- Helper AL. *Women in Labor: Mothers, Medicine, and Occupational Health in the United States, 1890–1980.* Columbus: Ohio State University Press; 2000.
- Hunter D. *Diseases of Occupations.* 6th ed. London: Blackwell; 1978. [The most complete edition for historical material.]
- La Berge AF. *Mission and Method: The Early Nineteenth-Century French Public Health Movement.* Cambridge: Cambridge University Press; 1992.
- LaDou J. Occupational and environmental medicine in the United States: A proposal to abolish workers' compensation and reestablish the public health model. *Int J Occup Environ Health.* 2006;12(2):154–168.
- McCord C. *A Blind Hog's Acorns: Vignettes of the Maladies of Workers.* Caldwell, NJ: The Blackburn Press; 2001. [Reissue of the first edition, published by Cloud, Inc., 1945.]
- Mock HE. *Industrial Medicine and Surgery.* Philadelphia: WB Saunders; 1919. [A masterpiece but virtually unavailable.]
- Selleck HB. *Occupational Health in America.* Detroit: Wayne State University Press; 1962. [The most comprehensive history of its type but virtually unavailable.]
- Starr P. *The Social Transformation of American Medicine.* New York: Basic Books; 1982:198–232.

## 2 TOXICOLOGY

Toxicology is a foundational science of occupational medicine that, together with epidemiology, must be understood in detail to be effective at any substantial level of responsibility in the field. Occupational and environmental physicians are often called upon to evaluate cases of suspected toxicity. They are even more often called upon to deal with complicated issues in toxicology related to risk assessment, regulation, and prevention.

This chapter is not intended to be a comprehensive introduction to toxicology but strives for a deep understanding in selected areas of toxicology that are most relevant. In this chapter, selected aspects of toxicology will be discussed as they apply to occupational and environmental medicine (OEM).

The principles of toxicology are familiar to physicians through analogy with pharmacology. What happens to drugs or other chemicals in the body, and the movement of drugs as reflected by their blood concentration, is called “pharmacokinetics” in pharmacology and “toxicokinetics” in toxicology. What the drugs or other chemicals do to the body, and how their effects relate to the magnitude of exposure, is called “pharmacodynamics” in pharmacology and “toxicodynamics” in toxicology. The principles are the same, however, and

most are also shared with nutrition science. That is why toxicologists prefer to use a general term for all chemicals foreign to the body, “xenobiotics,” which covers drugs, potentially toxic chemicals, nutrients, and even substances that act primarily as foreign bodies.

Some chemicals that are usually thought of as xenobiotics have endogenous sources in the body that produce low but detectable levels. For example, carbon monoxide is a product of the breakdown of bile acids. Ethylene is a product of the degradation of S-hydroxyethyl-mercapturic acid. Isoprene is produced during cholesterol synthesis. Some volatile organic compounds are produced by the conversion of tryptophan to indole. A few chlorinated organic compounds are produced by the oxidation of chloride. Nitric oxide, carbon monoxide, and even hydrogen sulfide have or may have physiological roles as neurotransmitters or neuromodulators.

## **TOXICOLOGY**

Toxicology is often called “the science of poisons,” but it is more accurately described as the science concerned with the behavior of chemicals in the body and the body’s response. (Some scientists include radiation, but in this book we do not.) As a discipline it draws most heavily from physiology, chemistry, biochemistry, pathology, pharmacology, and genomics. Toxicology is much more than the science of poisons because its scope has expanded since its modern inception to include identification of the limits of safety of chemical agents whether the chemicals are encountered in the workplace, the ambient environment, the home, or as pharmaceuticals, foods, food additives, cosmetics, or consumer products. These limits imply a socially determined level of acceptable risk and so toxicology has developed an important social science component in addition to its technical, biomedical aspect. Because the mechanisms of toxic effect are modified by adaptive processes, toxicology draws heavily on basic biomedical sciences. Table 2.1 outlines recognized applications and subdisciplines of toxicology. Most have their own professional bodies, societies, and sections within the Society of Toxicology and other organizations.

**Table 2.1.** Scope of Toxicology

Field of Specialization in Toxicology	Interdisciplinary with:
Clinical toxicology	
Drug adverse effects	Medicine, pharmacy, pediatrics, psychiatry
Drug abuse	Emergency medicine, pharmacy
Natural products (venoms, toxins)	Pharmacology, pharmacy
Suicide or accident prevention	Forensic medicine, pathology
Environmental and occupational toxicology	
Occupational toxicology	Occupational medicine, epidemiology
Environmental (media) toxicology	Environmental health epidemiology
Risk assessment	Political science, economics, law, public policy, epidemiology
Dosimetry and biological monitoring	Industrial hygiene, epidemiology
New product testing and product safety (toxicity screening)	Chemical engineering, cosmetology, food science, business, law, genetics, consumer protection, pharmaceuticals
Basic toxicology	
Toxicokinetics, receptor activation	Pharmacology
Metabolism of xenobiotics	Biochemistry, organic chemistry
Detection of xenobiotics	Medicine, physiology, pharmacology, biochemistry, psychology
Analytical toxicology	Clinical pathology
Target organ toxicology*	
Dermatotoxicology, ocular toxicology	Dermatology, ophthalmology, personal care, products testing
Inhalation toxicology	Pulmonary medicine
Hematotoxicology	Hematology, oncology

*(Continued)*

**Table 2.1.** (Continued)

Field of Specialization in Toxicology	Interdisciplinary with:
Hepatotoxicology	Gastroenterology
Neurotoxicology	Neurosciences, autonomic pharmacology
Renal toxicology	Nephrology
Immunotoxicology	Immunology, oncology, infectious diseases
Reproductive toxicology	Gynecology, urology, endocrinology, pediatrics, genetics
Carcinogenicity and genotoxicity	Oncology, internal medicine, genetics
Cardiovascular toxicology	Cardiology, neurosciences, pulmonary medicine, hypertension research
Toxicology pathology	Pathology, forensic, toxicity screening studies

\* Following terminology of the Society of Toxicology.

## Toxicological Concepts

Toxicology has had a long and colorful history. In ancient times, and today in some aboriginal cultures, poisons were used on arrows for hunting and warfare. Ancient texts, such as the Egyptian *Ebers Papyrus* of approximately 1500 BCE, show a sophisticated knowledge at the time of the toxicology of metals, opium, hemlock, and cyanogenic glycosides. Nicander of Colophon described the results of human experimentation with poisons in condemned criminals, and Galen of Pergamum, the great Hellenic physician, conducted experiments on animals. Toxicology was therefore one of the first sciences to develop a substantial empirical and experimental body of knowledge.

In 82 BCE, the Romans passed the first law against intentional poisoning (*Lex Cornelia*), which was later applied to adulterated foods. This was the beginning of regulation of chemical hazards. The Romans, however, overlooked the toxicity of lead (which was known to the Egyptians)

and saturated their society with it, from their food (lead acetate was used as a sweetener) to their plumbing (the word comes from *plumbum*, meaning “lead” in Latin) and industry, including smelting.

Toxicology developed further as part of forensic medicine in ancient times and through the Middle Ages in Europe. The study of poisons was advanced because of the need to identify their use in murders and to manage their use as ordeal poisons. Ordeal poisons were used in ancient and medieval times in the interrogation of witnesses, who were forced to consume a poison (often a plant containing physostigmine). If they lived through the overdose, they were presumed to have been telling the truth. Poisons were also used as methods of execution, most famously in the case of Socrates, who was dispatched with a bowl of hemlock. Criminals, including mass murderers, experimented on their victims in order to improve their nefarious techniques. (The record for individual murders is probably held by Catherine Deshayes, known as “La Voisin,” who killed over 2,000 victims in Paris during the reign of Louis XIV in the late seventeenth century.) Poisoning was also a tool of statecraft and insurrection through the assassination of political figures and of rivals to the throne. (This was a favorite ploy of the Medici and Borgia families in Italy.) Slaves and later servants were employed as food tasters in the homes of anyone royal, rich, or with political power. In 1818 the Spanish-born French medical scientist Mathieu Orfila established the first modern laboratory dedicated to forensic toxicology. The strong link between toxicology and forensic science continues to this day.

Antidotes to poisons and protective potions were desperately sought by rulers, rich merchants, and other people who had reason to believe that others wanted them out of the way by toxicological means. These antidotes and prophylactic potions were called, respectively, “theriac” or “mithridatum” and consisted mostly of combinations of herbs and secret ingredients. The formulation of theriacs, like the transmutation of base metals into gold, consumed much effort of alchemists and diverted the creativity of early proto-scientists and physicians. The literature of the time often uses potions as a plot device, as in *Romeo and Juliet*.

By the late Middle Ages in Europe, most of the essential elements and concepts necessary to make toxicology a modern science were in place: the experimental tradition (human and animal), a grounding in chemistry (thanks to alchemy), the concept of an antidote (theriac), the concept of bioavailability (contributed by Maimonides, a physician who was a leader in evidence-based medicine as well as religion and philosophy), the concept of tolerance or habituation (a feature of mithridatum), and even regulatory policy (the Lex Cornelia). However, there was an insurmountable obstacle to progress that was conceptual rather than technical.

Poisons were considered to be matter with magical properties, different from ordinary matter. Alchemists as well as ordinary people thought of a poison as a special type of substance different from the stuff of the rest of the world. They assumed, for example, that its magical powers could propagate beyond the vial or bottle of the poison and that poisons did not follow the ordinary rules of nature.

The breakthrough came with the insight of an extraordinary individual, one of history's great minds: Paracelsus (1493–1541). Paracelsus (Figure 2.1) was an itinerant scholar, teacher, and physician-alchemist who worked in Switzerland and Austria in the early sixteenth century. He had a mediocre education and a thoroughly disagreeable personality. He often got into trouble for being right. As a physician, he advocated mercury-based medications for syphilis, which worked, but the side effects caused problems following accusations that he poisoned his patients. He was passionately committed to empirical science, experimentation, and the advancement of alchemy, which he brought into modernity as toxicology. To Paracelsus the field owes the concept of the dose-response relationship ("... the dose makes the poison . . ."), the recognition that poisons tend to have specific and often characteristic syndromes of response (known today as "toxicodromes"), the concept of the therapeutic index (the margin in dose between effectiveness of a drug and its toxicity), and, above all, the recognition that poisons are not a special form of matter, that magic has nothing to do with it, and that anything that can be taken into the body is potentially toxic if the dose is high enough. ("All substances are poisons," he wrote.)



**Figure 2.1.** Paracelsus, a transformational figure in the history of toxicology. Ferguson Collection, University of Glasgow.

After Paracelsus, toxicology became, in effect, a subdiscipline of pharmacology and physiology. The mechanisms of drug effects (many of the drugs having been derived from classical toxins) were often elucidated through substances recognized to have toxic properties. For example, nicotine, atropine, muscarine, and physostigmine were used by pharmacologists and physiologists to investigate the autonomic nervous system. Chemical probes became an important tool in mainstream biomedical research.

Organic chemistry began to develop around 1800, and in subsequent years new chemicals (especially dyes and colorants) and synthetic materials were introduced. These were incorporated into processes and products during the Industrial Revolution, resulting in severe chemical contamination in and around smelters, factories, and home workshops. Food adulteration and unsafe products became serious public health issues. New medicines resulted in new side effects. The lessons of toxicology and forensic medicine became increasingly valuable commercially and legally as poisoning cases became more complicated. Finally, with the passage of the landmark Public Health Act of 1848 by the British Parliament, and its subsequent amendments and associated acts, a new consensus for regulation and the protection of public health was achieved. The new regulatory regime required new ways of thinking and tools for detection and measurement, which further advanced toxicological sciences. By 1906 the uproar in the United States that followed revelations about food safety in Upton Sinclair's muckraking book *The Jungle* led to the first Food and Drug Act and, much later, in 1955 the Food and Drug Administration. (Sinclair was disappointed, however, because the book was mostly about labor exploitation and occupational hazards and little was done about that.)

In the twentieth century, momentum for the development of toxicology as an independent discipline came (in roughly chronological order) from food safety, chemical warfare during World War I and the threat of it in World War II, product safety (especially cosmetics and food additives but also industrial chemicals), radiation biology (a close sister discipline to toxicology), pesticide research, concern for environmental quality, occupational medicine, recent refinements in methodology of epidemiology and risk assessment, materials science and biocompatibility, molecular genetics and carcinogenesis research, advances in immunology, and regulatory needs to support environmental protection. This broad sweep was punctuated by periodic tragedies that heightened awareness. During Prohibition in the United States, for example, there were many outbreaks of toxicity due to methanol and lead from moonshine, and of a tabes-like

dorsal root neuropathy (“Ginger Jake paralysis”) due to massive consumption of an alcohol-containing patent medicine adulterated with a plasticizer, tri-ortho-cresyl phosphate. In 1938 a new and legal formulation of the antibiotic sulfanilamide dissolved it in diethylene glycol, which is nephrotoxic, resulting in the deaths of over 100 people and the recognition that even legitimate medications could be unsafe apart from their predictable side effects.

Contemporary public attitudes toward toxicology are shaped by regulatory philosophy, the risk-averse attitudes of society, and concern that environmental pollution levels are increasing. The public also expects an explanation for diseases and assurance of safety. In particular, the public tends to revert to older, even ancient, ways of thinking about toxicology when faced with a threat. For example, many, if not most, people tend to think of chemicals as either “toxic” or “non-toxic” rather than falling on a continuum of toxicity depending on dose. People are often resistant to the idea that if a toxic substance is contained or if there is no pathway of exposure to reach people, there is no risk. These are echoes of belief in the magical properties of poisons. These ways of thinking are deeply embedded in risk perception and the public’s way of interpreting chemical hazards, which are factors to be dealt with in risk communication (see Chapter 7).

## Occupational Toxicology

From the very beginning, the disciplines of occupational medicine and toxicology have been inextricably linked. There are allusions to occupational diseases in ancient texts. The first clinically accurate descriptions of occupational disorders, miners’ phthisis (silicosis complicated by other prevalent lung diseases) and mercury poisoning, were contributed by Ulrich Ellenbogen, Georgius Agricola, and Paracelsus, all before 1600. Observations of great toxicological significance were made by all the early pioneers of occupational medicine, such as Bernardino Ramazzini, but the contributions of Charles Turner Thackrah to understanding lead toxicity stand out among clinical observations before 1850. Many of these observations led to

important lines of investigation in basic toxicology: in 1775 the surgeon Percival Pott correctly associated cancer of the scrotum in chimney sweeps with exposure to soot in conditions of poor hygiene, an observation that led directly, if belatedly, to the work of Katsusaburo Yamagiwa and Koichi Ichikawa in 1912 demonstrating that coal tar products induce skin cancer when applied to experimental animals, which led thereafter to modern carcinogen bioassays and contemporary theories of chemical carcinogenesis. As the discipline of occupational medicine developed, such leading figures as Alice Hamilton and Carey McCord grounded their interests and expertise primarily in toxicology.

As workplace conditions have improved historically, the detection of occupational disorders has become a more subtle undertaking, requiring increasingly sophisticated methods contributed by toxicologists and epidemiologists. As the presentation of occupational disorders has become less dramatic, particularly in the case of toxic hazards, the standard diagnostic methods of clinical medicine have become less satisfactory and have had to be supplemented—and largely replaced in practice—by laboratory methods for early detection of excessive exposure and for diagnosis of often nonspecific clinical presentations as toxic in origin.

Occupational medicine and the scientific discipline of toxicology itself have been on intertwined paths for many years. From its early preoccupation with particularly toxic chemicals, from which it gained its essential definition as the science of poisons, toxicology has expanded its scope to include biological mechanisms of toxicity and host defenses (or “resistance”) against toxicity. In recent years, toxicology has become highly specialized in the area of risk definition, identifying the level of hazard peculiar to a particular chemical exposure and the limits of acceptably safe exposure. These issues go far beyond characterizing the effects of poisons because most of the chemicals of modern concern are not classically “poisons” in the sense of being potentially lethal at low doses. These chemicals include those encountered in the workplace, the ambient environment, and the home, as well as pharmaceuticals, food constituents, food additives, cosmetics,

and consumer products. Because the delineation of safe levels of exposure assumes a socially determined level of acceptable risk (implicit in the definition of “safety”), toxicology has, along with occupational medicine, become increasingly influenced in both theory and practice by the behavioral sciences and by regulatory bodies.

However, the relationship between toxicology and occupational medicine is not all one-way, with toxicology providing the underpinning of occupational medicine practice. The scientific communication flows in the opposite direction as well. Occupational exposures are among the few socially acceptable opportunities toxicologists have for studying the effects of chemical exposures on human beings at nontrivial exposure levels. Well-conducted occupational studies are therefore invaluable in providing important information to toxicologists on the human response to exposure to many chemicals. The methodology for such studies is usually epidemiological because relatively large populations are studied. However, individual cases and case series of workers who have developed a toxic response to a chemical exposure are uniquely valuable in occupational medicine and toxicology, although they are somewhat out of favor in other areas of medicine. Such case reports reflect the rarity of some exposures and some diseases (e.g., hepatic angiosarcoma) and provide clues to mechanisms of toxicity and defense. They also reflect a collective sharing of experience in treatment that can be obtained in no other way, because clinically obvious exposures typically occur infrequently and unpredictably and do not lend themselves to clinical trials or formal studies. Thus, occupational medicine provides the science of toxicology with much in return.

## TOXICOKINETICS

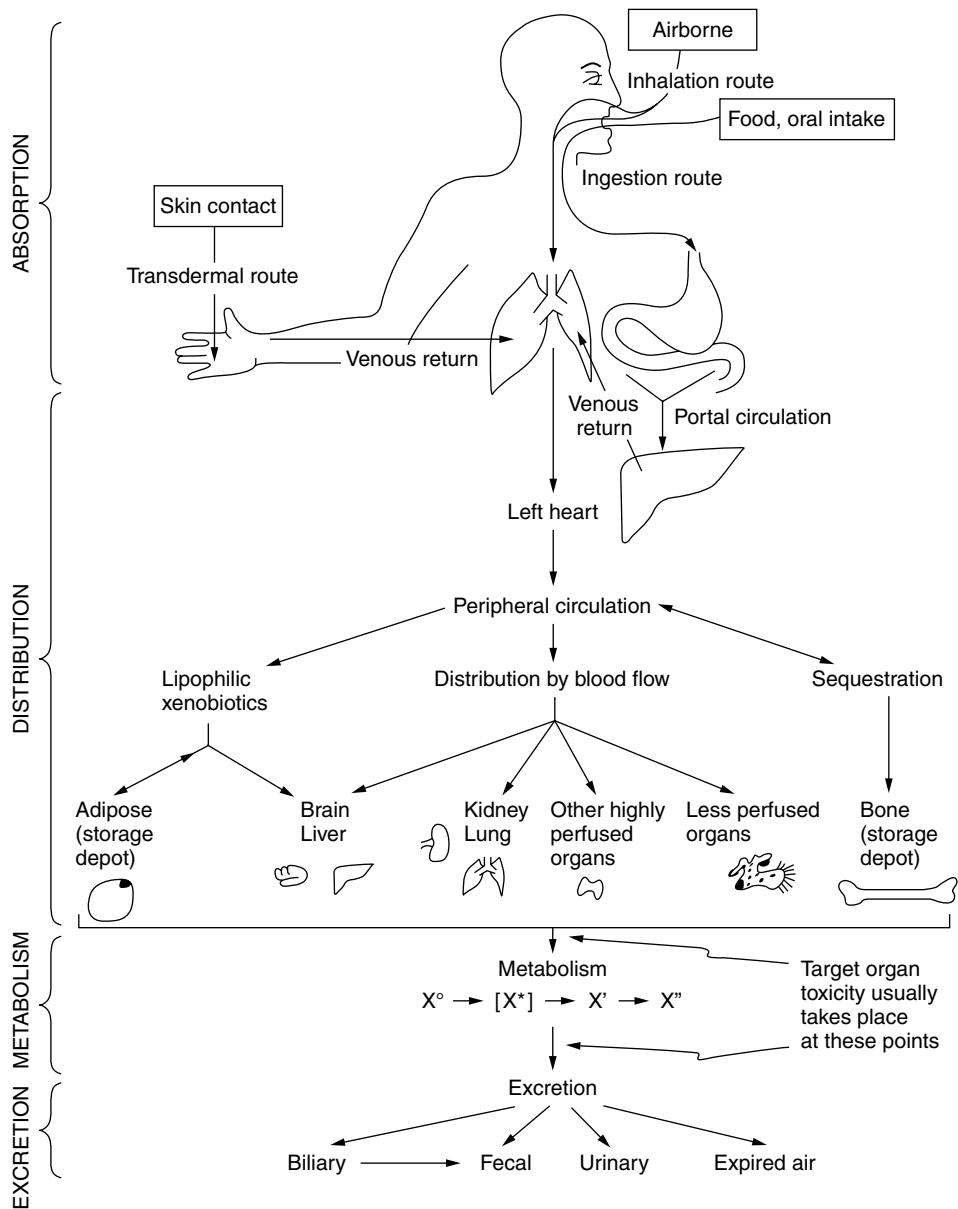
For convenience in terminology, we may refer to all substances not normally present in the body and introduced from outside as “xenobiotics” (from the Greek *xenos*, meaning “foreign”). Xenobiotics may be drugs, food constituents, occupational chemical exposures, or environmental chemical exposures. Regardless of their effect or

origin, the behavior of xenobiotics in the body can be described by general terms and models reflecting the mechanisms by which exposure occurs and by which the body handles the chemical. From the standpoint of evolutionary biology, it is supposed that these mechanisms developed in response to selection pressures reflecting two biological needs: to detoxify and excrete harmful substances ingested in foods (especially in spoiled or putrefied foodstuffs) and to metabolize endogenous chemical compounds (such as steroid hormones).

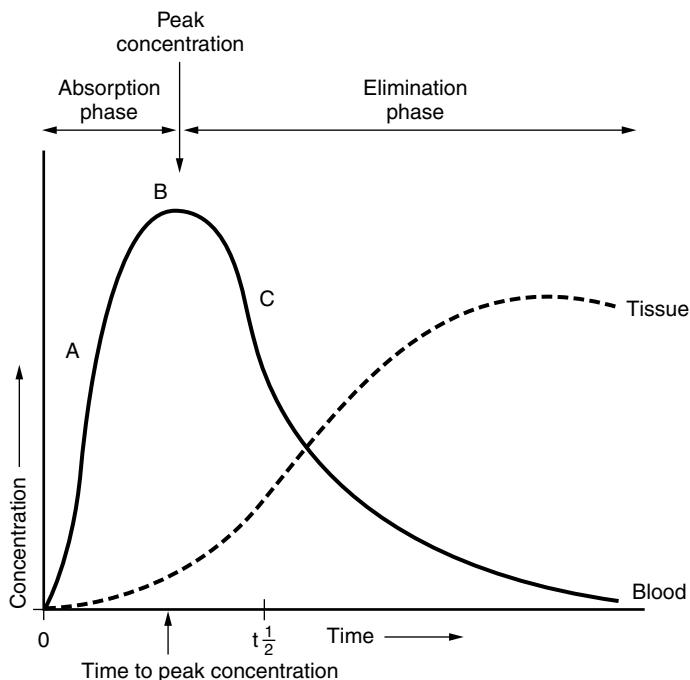
Toxicokinetics is the toxicological analogy to pharmacokinetics and is based on the same concepts. It is therefore often useful to remember the disposition and metabolism of common drugs in thinking through the behavior of a toxic chemical or other xenobiotics. Four terms describe the disposition of xenobiotics: “absorption,” “distribution,” “metabolism,” and “excretion.” Figure 2.2 presents these concepts in schematic form, as they apply to occupational medicine. Modeled together, the terms describe the entry, local and overall accumulation, transformation, and removal from the body of the xenobiotic. Because tissue levels depend on transport of the xenobiotic to the target organ and the degree to which the xenobiotic partitions or is sequestered into the tissue, the kinetics of the xenobiotic determine the delivery of the xenobiotic to the target organ at the receptor level, where the toxic effect occurs.

Figure 2.3 presents an idealized curve for the plasma concentration of a xenobiotic. Such curves describe the behavior of a single dose, or bolus, of a xenobiotic in the body. In practice, the curves are typically presented in logarithmic transformation so that they can be plotted as straight lines.

Xenobiotics encounter the body through contact with a medium, which may be air, water, food, injection, implantation, a liquid in physical contact with skin, a solid in physical contact with skin, a consumer product, and so forth. How the xenobiotic gets there is the basic problem of exposure assessment (Chapter 4), as well as studies of the fate and disposition of chemicals in the environment.



**Figure 2.2.** Overview of toxicokinetics. Adapted from Guidotti TL, "Principles of Occupational Toxicology" in *Occupational Medicine: Principles and Practical Applications*, 3rd ed., edited by Zeng C, Nickerson B, and Hovath EP, pp. 70–84, Copyright © Mosby/Elsevier (1994).



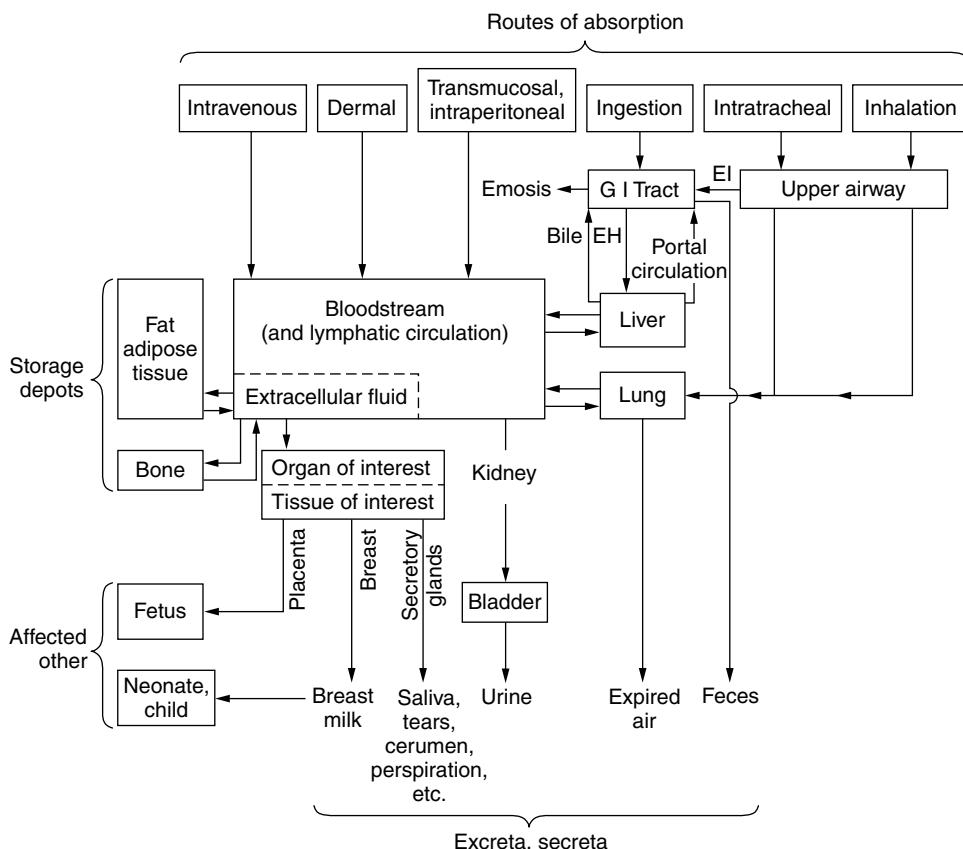
**Figure A.** Idealized kinetic curves for a xenobiotic introduced into the body.

- A absorption, entry into bloodstream
- B peak blood concentration
- C distribution to tissue

**Figure 2.3.** Idealized curve for the plasma concentration of a xenobiotic, administered intravenously or by inhalation.

Figure 2.4 presents a general model for the behavior of a xenobiotic in the body and provides a framework for tracing the path of an agent. An “exposure,” in toxicology, can be what the body encounters or what the body takes up. How much is taken up depends on how much of the xenobiotic reaches the body, how much of it is available to the body once it gets there, how much is absorbed by or enters into the body, and how much enters the circulation, is mobilized, and is distributed.

The degree to which a xenobiotic is absorbed in the form in which it is encountered is called “bioavailability,” and depends on many physical and some chemical properties: solubility (in water and lipid), ionization, particle size (volume and surface area), temperature,



**Figure 2.4.** Principal natural routes of absorption, distribution, excretion, and deposition of xenobiotics (EI = expectoration and ingestion; EH = enterohepatic circulation). Adapted from Guidotti TL. Figure appearing in *Assessment for Environmental Health* by Mark Robson and William Toscano, Jossey Bass, 2007. Copyright © 2007 by John Wiley & Sons Inc. All Rights Reserved. Reprinted with permission.

emulsification, the presence of solvents, the presence of binding agents, and many other factors. Bioavailability is most familiar in the context of drugs but is also meaningful in occupational and environmental toxicology. For example, the degree to which lead is bioavailable to the body depends on its form. Lead salts, such as lead carbonate, are highly soluble and readily absorbed. Lead chloride (which is not a salt) is not. Metallic lead is poorly soluble and rarely causes lead poisoning unless particles of lead, like bullet fragments, are lodged in highly vascular tissue or inhaled. Then, the degree of

absorption depends on particle size. Smaller particles have much larger surface areas proportionate to their volume and, because absorption occurs from the surface, metallic lead is more bioavailable in smaller particles.

## Absorption

Xenobiotics may enter the body through any of the “portals,” or routes of entry, listed in Table 2.2. The rate of absorption is the most important determinant of the peak levels that will be reached in plasma. For many toxic substances, this is the prime determinant of acute toxicity.

Transdermal, inhalation, and transmucosal exposures represent the most common routes of human exposure in the workplace, whereas ingestion, inhalation, and transdermal exposures are the most common routes in environmental exposure. Transdermal, transmucosal, ocular, and ingestional routes are also particularly important in cosmetics and consumer safety. Intramuscular, intradermal, subcutaneous, and intravenous routes are confined almost entirely to medical administration of drugs. The experimental study of xenobiotics usually involves exposure by inhalation, ingestion, or the intraperitoneal route to determine toxicity, and intravenous administration to determine the toxicokinetics beyond the absorption phase; special applications may require special techniques, however, such as intrapleural administration to study carcinogenic potential of mineral fibers in producing mesothelioma, or intratracheal instillation as an alternative to inhalation.

In occupational and environmental medicine, by far the most important routes of exposure are skin contact and inhalation. The most significant portals of entry are therefore absorption through skin and inhalation. Ingestion, resulting from eating or placing objects such as cigarettes in the mouth in a situation where the object or the hands may have been contaminated, or in suicide attempts, is not a common problem in occupational medicine but appears from time to time and is highly significant in

**Table 2.2.** Routes of Absorption of Xenobiotics

Route of Exposure	Site of Entry	Context	Characteristics
Transdermal	Skin	Environmental, occupational, experimental	Favors lipophilic xenobiotics, may be premetabolism, slow
Transmucosal	Mucous membranes	Occupational, medical	Rapid entry into circulation
Ocular	Mucosal membranes of eye	Occupational (splash)	Local effects, access to circulation
Intramuscular	Muscle	Medical	Relatively slow release
Intradermal	Skin, deep	Medical	Relatively slow release
Subcutaneous	Skin, superficial	Medical	Relatively slow release
Intravenous	In vein	Medical	Immediate entry into circulation, mixing
Intra-arterial	In artery	Medical	Immediate entry into circulation, delivery of bolus to tissue downstream
Inhalation	Upper airway and lung	Environmental, occupational, medical	Rapid entry into circulation, local effects
Intratracheal	Trachea, lung	Experimental	Rapid delivery to lung
Ingestion	Gastrointestinal tract (stomach, small intestine)	Physiological	Relatively slow, rapid delivery to liver from gastrointestinal tract
Intraperitoneal, intrapleural	Peritoneal and pleural cavities	Medical, experimental	Rapid absorption, local effects

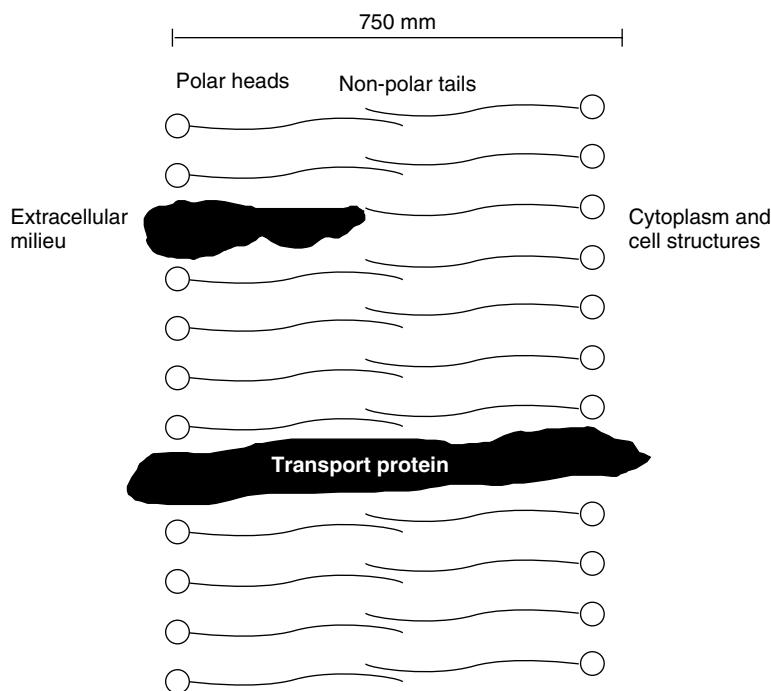
environmental toxicology and the toxicology of food safety and consumer products. Splashes into the eyes are more often associated with local eye irritation and only rarely with absorption and systemic toxicity. Other routes of exposure, such as intravenous infusion or implantation of soluble agents, are artificial and seldom seen outside of medical care and experimental studies but may be very important in those settings. Needlestick injuries may inject pathogens or drugs intended for a patient into a healthcare worker, for example.

Among the natural routes of entry, inhalation and transmucosal absorption of xenobiotics are the most direct and efficient routes into the bloodstream. The xenobiotic that passes across the alveolar membrane enters quickly into the bloodstream, rivaling intravenous injection for efficiency, and is rapidly mixed and distributed by the circulation throughout the body. On the other hand, xenobiotics that are absorbed through the skin tend to move more slowly and may be subject to metabolism in the cells of the skin. Ingestion, on the other hand, is highly variable and diverts the absorbed xenobiotic to the liver through the portal circulation. Ingested xenobiotics therefore see the liver first and are more likely to be metabolized before being distributed to the rest of the body. These routes of entry are not mutually exclusive. Inhalation of poorly soluble dusts such as silica, for example, may result in ingestion of the same material because of clearance from the lung bringing the material up the mucociliary escalator, where it is swallowed or expectorated.

The toxicity of the xenobiotic may or may not involve the organ of first contact or site of entry. For example, carbon monoxide enters the body by inhalation but has no appreciable toxic effect on the lung. Other chemicals, such as irritants applied to the skin, may cause local toxicity without significant absorption or effect on the rest of the body.

The rate at which a xenobiotic enters the bloodstream is determined by absorption across the barrier presented by the given route

of exposure, particularly the lipid bilayer of the living cell membrane. The cell membrane is not just a lipid bilayer consisting of phospholipids, cholesterol, and sphingolipids (including phosphatidyl choline) that orient their heads toward the extracellular milieu and cytosol and their tails in the bilayer. It is studded with pores, carrier proteins, and active channels to facilitate the entry of some nutrient or electrolytic chemicals and the exclusion of others (see Figure 2.5). Absorption of xenobiotics across membranes is determined for the most part by the chemical and physical properties of the agent and its surrounding chemical environment. In general, lipid-soluble (lipophilic, or hydrophobic) substances are absorbed



**Figure 2.5.** The lipid bilayer of cell membranes.

more readily than water-soluble substances (hydrophilic) across lipid-rich barriers such as skin. This can be tested in the laboratory by determining how the xenobiotic partitions between octanol, an organic solvent, and water—a standard test for predicting absorption. Some hydrophilicity is required, however, to get out of the other side of the lipid bilayer.

Another, related characteristic determining absorption is the ionization of the xenobiotic. Ionized molecules have a harder time crossing the membrane than nonionized molecules. Therefore a weak base (which includes many medications, nicotine, caffeine, and a few occupationally significant toxic chemicals) will cross a cell membrane more efficiently in an alkaline environment, such as the gut, and a weak acid (of which there are many, including salicylic acid) will cross more readily in an acidic environment, such as the stomach or the surface of the skin. The fraction of a weak base or weak acid that is ionized can be calculated from the equilibrium equation for the dissociation constant  $K$  and the pH. The  $pK_a$  ( $-\log K$ ) is a useful parameter that is provided in many toxicological resources for weak acids and bases. When the  $pK_a$  equals the pH, 50 percent of the weak acid or base is ionized.

The rate at which a xenobiotic enters the bloodstream is the most important determinant of the peak levels that will be reached in the bloodstream. Indeed, for many toxic substances, this is the prime determinant of acute toxicity. The most rapid entry into the bloodstream and therefore the highest peak level is accomplished by intravenous injection, but this route is limited to medical and experimental applications. However, intravenous administration of a given amount of the xenobiotic permits direct study of the agent's distribution, metabolism, excretion, and acute toxicity without the complication of absorption factors modifying its rate of uptake. A known dose can be injected and its concentration in the blood studied without having to take into account slow or sustained absorption. Inhalation of soluble xenobiotics also results in rapid absorption into the bloodstream, because in the lung, the agent is brought into very close contact with the circulation at the alveolar level.

### ***Gastrointestinal Absorption***

Ingested xenobiotics pass across mucosal tissues in the mouth, where significant absorption can take place, and pass through very different regimes of pH in the lower gastrointestinal tract. In the stomach, the pH is generally low, so weak acids will be in their nonionized form and therefore more efficiently absorbed. Weak bases will be ionized and so will be poorly absorbed. In the gut, the much higher pH favors the absorption of weak bases and poor absorption of weak acids. There are also a number of special transport mechanisms, including one for glucose that absorbs chemicals that resemble that sugar, and for iron, which may take up lead instead. For example, strychnine, which was once a popular rat poison, is a weak base. After it is ingested, nothing much happens at first because, being almost completely ionized, it is poorly absorbed in the stomach. An hour or so later, the patient is suddenly overcome with excruciating pain from tetanic seizures (due to competitive antagonism of an inhibitory neurotransmitter that relaxes muscles after contraction). The reason is that the weak base passes into the small intestine, where it is efficiently absorbed, with devastating results.

This general pattern of gastrointestinal absorption is not absolute. There may be departures from predicted behavior in absorption that reflect delayed gastric emptying and gut motility. Bile acids render hydrophobic compounds more soluble and more finely emulsified and therefore more easily absorbed beyond the duodenum. Large molecules are absorbed in the gut more efficiently than predicted, presumably by endocytosis. Some hydrolysis takes place in the lumen of the gut, especially of polypeptides, further enhancing their absorption.

Ingestion is the most common route for suicidal or accidental intake of toxic chemicals or drugs and is more important in clinical toxicology as practiced in the emergency room than in occupational toxicology. Nevertheless, in occupational settings ingestion may occur in many situations: contaminated food and eating areas,

release from objects held in the mouth, consumption of chemicals by accident (usually in unlabeled bottles), attempted suicide, and as a secondary route of exposure occurring indirectly as a result of inhalation. In the last-mentioned case, particles deposited in the lung may be coughed up and then swallowed, entering the gastrointestinal tract. Ingestion results in slower absorption of the agent than inhalation and transmucosal absorption, but is usually much faster than transdermal exposure because absorption often begins in the capillary bed of the stomach.

### ***Inhalation***

Exposure by inhalation results in relatively efficient absorption of gases if the gas can penetrate to the alveolar-capillary bed. Whether the gas will penetrate efficiently depends on its solubility in water, reflecting clearance rates in the bronchial tree. Once having penetrated to the alveolar level, however, gases are readily absorbed across the alveolar-capillary bed by simple passive diffusion. Absorption across the alveolar membrane in the lung is usually very efficient and complete, and entry into the bloodstream is limited only by the xenobiotic's solubility in plasma, an aqueous medium. Particles, on the other hand, are subject to a number of host defense mechanisms in the respiratory tract that limit the efficiency of penetration to the alveolar level to relatively small particles. Once there, their size prevents them from passing directly into the bloodstream and they must dissolve or be digested by macrophages before their constituent chemical contents can be absorbed and enter the bloodstream. Particles may contribute to systemic toxicity if they are composed of a soluble material such as lead or polycyclic aromatic hydrocarbons. For this reason, inhalation of toxic gases is usually associated with acute systemic toxicity or vascular injury to the lung (resulting in pulmonary edema), but particle deposition in the lung is usually associated with localized pulmonary effects and chronic systemic toxicity. This topic will be revisited in greater detail in the section on inhalation toxicology.

### *Skin*

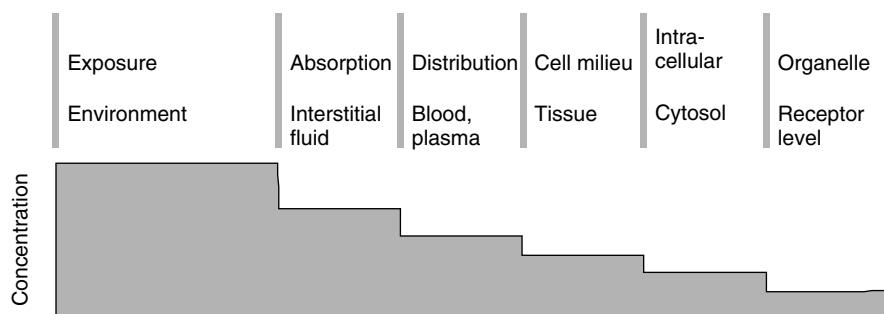
Although the skin acts as a barrier to the entry of environmental agents into the body, it is sufficiently permeable to be a major route of entry into the body for many chemicals, particularly those that are readily lipid-soluble. Absorption across the skin is highly variable, depending on skin characteristics and the solubility of the xenobiotic in fat. Most transdermal absorption occurs directly across the superficial layers of the skin, or the stratum corneum, which consists of nonliving, keratinized cells, and the other living cell layers of the epidermis to be absorbed in the capillary bed of the dermis. Some chemicals applied to the skin may gain entry through a shortcut, passing more rapidly through hair follicles and sebaceous gland ducts. When the skin is injured with open wounds or abrasion, or in the presence of a skin rash, absorption across the skin is much faster.

In occupational medicine, transcutaneous absorption is a particular problem in the toxicology of pesticides, solvents, and halogenated hydrocarbons generally. Some agents may be significantly metabolized by enzyme systems in the skin, but most appear to gain entry into the bloodstream unchanged. Rates of penetration are known for many chemicals from experimental studies using pig skin as a barrier. The thickness of the skin is a major factor, as is the temperature, humidity, blood flow to the skin, and the presence of solvents or surfactants, all of which facilitate passage. When the skin is injured through excoriation, open wounds, abrasion of the epidermis, or the presence of a skin rash, absorption is much faster. The skin of the forearm is used as the standard in human experiments because it is intermediate between sites of low penetration (the palm and the plantar surface of the foot) and high penetration (back and scalp). Mucous membranes absorb xenobiotics much more rapidly than skin. In occupational medicine, the skin is also of major importance as a route of entry, particularly for pesticides, solvents, and halogenated hydrocarbons.

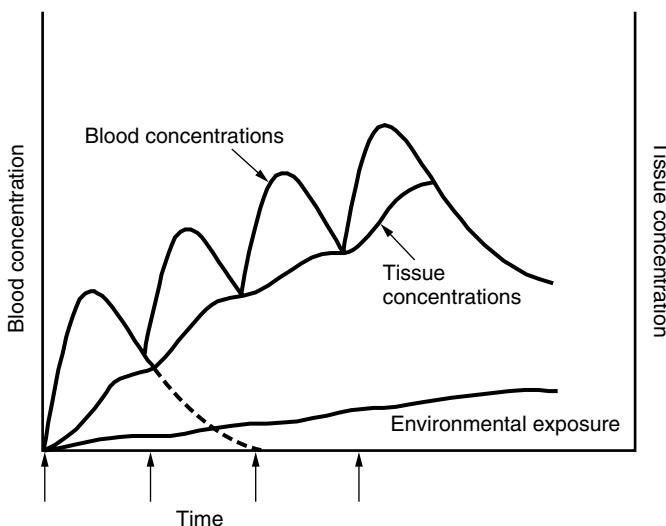
## Distribution

Once the xenobiotic is absorbed and enters the bloodstream, it is transported to the capillary level in tissues of the body, where it becomes available for uptake by the target organ. After one pass through the circulation, the xenobiotic is uniformly mixed in arterial blood regardless of its route of entry. It is then carried to tissue in blood or extracellular fluid, and once it reaches the site, it is taken up by the tissue, depending on affinity, which is largely governed by the same principles as those previously described in the absorption phase. The xenobiotic may enter the cell or interact with receptors on the surface, or it may pass into cellular organelles (such as the mitochondria). At each stage in this journey the concentration diminishes, but it is at the receptor level, on or in the cell, that the concentration really matters (see Figure 2.6).

When a bolus is absorbed, as by intravenous administration or inhalation of a gas, the peripheral tissues are presented with a rapidly increasing concentration in the blood that peaks and then declines as the xenobiotic is distributed to tissues throughout the body and is then removed by metabolism, excretion, or storage. This is illustrated in Figure 2.7. Multiple doses of a drug result in a staircase-like increase in blood and tissue concentration until a steady state is attained, an important concept in pharmacokinetics. The typical exposure pattern in occupational and environmental medicine, on the other hand, is much less distinct because the cumulative exposure takes place over a longer time and at lower levels, thus attenuating the exposure.



**Figure 2.6.** Plasma and tissue levels of a xenobiotic after absorption.



**Figure 2.7.** Cumulative circulating blood levels of a xenobiotic after multiple exposures.

When a xenobiotic is dissolved in plasma, some fraction of the total usually binds to circulating proteins, particularly albumin. Binding occurs quickly, and an equilibrium is established between the fraction of the xenobiotic bound to a plasma protein, which cannot leave the vascular space, and that dissolved in the plasma, which is free to diffuse or be taken up by tissues. As the concentration of free xenobiotic falls in plasma, some molecules will separate from their binding sites until a new equilibrium is reached. Binding proteins therefore act as a storage and release mechanism, maintaining a more even plasma concentration than would otherwise be the case and reducing the peak concentration that would otherwise be delivered to tissues. For example, DDT, dieldrin, and some other organochlorines are more than 99.9 percent bound (and carried by both albumin and LDL in the circulation). Parathion is 99 percent bound, mostly to albumin. Nicotine, however, is only 25 percent bound, mostly to albumin. Binding in plasma limits the volume of distribution to the vascular space, keeps many lipophilic xenobiotics in circulation, represents a major storage depot for some water-soluble xenobiotics, and allows mobilization and redistribution over time. Bound xenobiotics may be

displaced by other xenobiotics, a well-known mechanism for drug interactions. As a practical matter, this is of greatest significance in drug-related toxicology as a mechanism of drug interaction and overdose and is seldom a consideration in occupational toxicology.

In most cases, delivery of a xenobiotic depends on the blood supply to the tissue relative to its weight. When the xenobiotic is neither particularly lipophilic nor sequestered nor preferentially taken up by some organ-specific mechanism, it is largely distributed on the basis of blood flow to the target organ; organs with greater perfusion will tend to accumulate the xenobiotic simply because of the increased total amount presented to them. The lung is the only organ of the body to receive 100 percent of the cardiac output at a tissue level. (The heart, functioning as a pump, moves blood in bulk but is itself nourished by a much smaller coronary artery system.) Not surprisingly, the lung is a principal target organ for blood-borne as well as airborne xenobiotics. The liver and kidneys each receive massive fractions of the cardiac output and are therefore presented with circulating xenobiotics in great quantities. The brain also receives a disproportionate fraction of the cardiac output but is partly protected by the blood-brain barrier; this barrier works well for most polar xenobiotics but is generally permeable to lipophilic compounds. Barriers to penetration within the body follow the same principles as barriers to absorption.

Uptake of a xenobiotic by an organ from the plasma also depends on the affinity of the tissue for the material. The same principles apply as in absorption into the body. When the xenobiotic is stored in a tissue, the blood concentration comes into an equilibrium when the xenobiotic is released from the tissue and may stay fairly constant for the remaining life of the individual.

Distribution of a xenobiotic from the bloodstream into the tissue depends mainly on the solubility of the xenobiotic in fat. Lipophilic agents will accumulate in adipose tissue or lipid-rich organs such as the nervous system or in the liver. Organs with a high adipose or lipid content accumulate much larger concentrations of highly lipophilic xenobiotics, such as the PCBs, than plasma or other organs. This is useful scientifically as a means of measuring body burden because

subcutaneous fat biopsies are easy to perform. When an obese individual who has a high level of fat-soluble toxic chemical stored in adipose tissue rapidly loses weight as a result of dieting, food deprivation, unaccustomed exercise, or cachexia, the xenobiotic may be mobilized, and a rapidly climbing circulating level of the agent may rise to toxic levels. In general, however, the principal significance of adipose and intracellular lipid is as a storage depot, in that the blood concentration comes into an equilibrium when the xenobiotic is released from the tissue in which it was stored, remaining fairly constant for the remaining life of the individual. The xenobiotic can rarely be effectively purged from the body in this situation because of the extent of the storage, although one can steadily reduce the body burden over time by vigorous removal from plasma to force mobilization. Accumulation of toxic agents in breast tissue, which is predominantly adipose, and subsequent excretion into breast milk is the major route of exposure to a variety of xenobiotics for newborns who breast-feed.

When the properties of the organ attract and bind metals, as in bone, a metal or semimetal will be sequestered and will tend to accumulate over time, making the organ serve as a storage depot. Metal ions that resemble calcium, such as lead or strontium, are readily taken up by the surface of bone or tooth enamel by diffusion and exchange, insinuating themselves into the hydroxyapatite crystal. Presumably some is also taken up by the osteoblast and incorporated into new bone. Likewise, fluoride ion can substitute for a hydroxy group in calcified tissue. This process results in storage of a large fraction of certain xenobiotics; for example, some 90 percent of lead may be sequestered in this fashion, giving rise to the familiar radiographic signs of lead poisoning in children. Bone is a metabolically active tissue, however, and sequestered xenobiotics can be mobilized by increased osteolytic activity, such as in remodeling. Metals can also be remobilized by the administration of chelating agents, which promote excretion and the establishment of a new equilibrium with plasma. This is rarely advisable as therapy, however, because overly vigorous mobilization can result in acute toxicity, nutrient depletion, and demineralization of bone.

Binding proteins in various organs sequester some xenobiotics in organs that are susceptible to injury. The proximal tubule of the kidney, for example, is rich in metallothionein and is a depot for storage of many xenobiotics, including cadmium, which is nephrotoxic. The liver is rich in ligandin and metallothionein as well as lipid. In the liver, the portal circulation also delivers ingested xenobiotics at high concentrations directly from the stomach and small intestine. This routing provides an opportunity for metabolism to take place before the xenobiotic enters the general circulation, and the liver is the principal metabolic site for xenobiotics, as it is for nutrients. Circulating xenobiotics, such as kepone, may even be taken up by the liver and reprocessed through biliary excretion and reabsorption through the enterohepatic circulation. This is called “enterohepatic circulation,” and it is an important mechanism by which some xenobiotics maintain their concentration in the circulation.

Entry into some tissues is restricted by special barriers to passage, such as the blood-brain barrier, the placenta (which is not an effective barrier and actually is more of a delivery system to the fetus), the blood-testis barrier (which impedes small molecules), and the blood-retina barrier (which impedes large molecules). These barriers define “compartments” in which the entrance, exit, and turnover of xenobiotics are different from the rest of the body, an important concept in toxicokinetic modeling.

The blood-brain barrier is formed by tight junctions of endothelial cells and by glial cells that protect nerve tissue. There is also a low protein content in cerebrospinal fluid, which reduces binding and therefore release of xenobiotics. The blood-brain barrier is incomplete at birth and allows passage of potentially toxic agents, whether xenobiotic or endogenous (the mechanism of kernicterus in newborns). However, the blood-brain barrier also diverges from what one would expect. It is remarkably effective at excluding dioxins and other organochlorines, despite their lipophilicity and small size. There are also specialized carriers and transport mechanisms. Energy-dependent “drug transporters” secrete certain chemicals out of the brain back into the circulation, complicating chemotherapy.

There is also a carrier that brings magnesium into the brain and may also carry similar metals such as manganese.

The familiar principles of pharmacology also apply to toxicology with respect to volume of distribution,  $V_d$ , and the kinetics of initial distribution and blood concentration. This concept is less useful in occupational and environmental toxicology with respect to initial concentration and distribution, because exposure is much more variable (being rarely a single, rapidly absorbed dose). However, it is central to modeling toxicokinetics, which will be discussed later.

## Metabolism

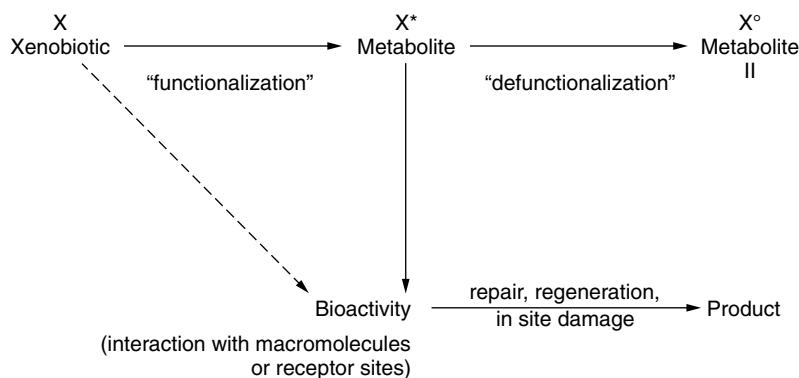
Many xenobiotics undergo chemical reactions in the body that affect their toxicity. These reactions are called “biotransformation” and are the result of metabolic pathways. An active xenobiotic may be transformed into an inactive metabolite, effectively removing the agent from the body. Alternatively, an inactive precursor may be transformed into an active metabolite, in effect incorporating a step between absorption and distribution when metabolism occurs in the liver after ingestion.

Most xenobiotics are substrates for intracellular enzyme systems, most of which appear to have evolved as mechanisms for clearing endogenous, mainly steroid, hormones or foreign substances taken in with food. In a series of steps, these enzyme systems transform the xenobiotic from the original compound into a series of stable metabolites, often through intermediate unstable compounds. For many xenobiotics there are multiple pathways of metabolism, resulting in numerous metabolites. These transformations may have the effect of either “detoxifying,” by rendering the agent toxicologically inactive, or of “activation,” by converting the native agent into a metabolite that is more active in producing the same or another toxic effect. An active xenobiotic may be transformed into an inactive metabolite, effectively removing the agent from the body in its toxicologically active form. In general, the enzyme systems that biotransform xenobiotics tend to convert nonpolar, lipid-soluble compounds

into polar, water-soluble products that are more easily excreted in urine or bile. The general pattern consists of two phases. Figure 2.8 illustrates the net effect of this transformation.

Phase I of the metabolic process involves altering the molecule by the addition or modification of functional chemical groups. This usually results in activation, especially in the very important “mixed function oxidase” (MFO) system. For example, it may result in a metabolite capable of interacting with macromolecules, such as DNA, in the earliest steps of the first stage of carcinogenesis. The mixed function oxidase system requires a great deal of metabolic energy and is closely linked with cytochrome oxidase electron transport, which provides it. Because the particular cytochrome most closely linked with the system has a spectral absorption peak at 450 nm, there is frequent reference in the literature to “P450” as an indicator of MFO activity. Phase I reactions include oxidations, reductions, hydrolysis, hydration, dehalogenation, and other reactions.

Phase II involves the removal or conversion of chemical groups or the attachment of more polar groups in such a way as to render the molecule as a whole more polar and hydrophilic, and therefore more easily excretable by the kidney (and less diffusible back across the renal tubular epithelium after filtration). In the process, the activated xeno-



**Figure 2.8.** Basic pattern of biotransformation of xenobiotics. Adapted from Guidotti TL, “Principles of Occupational Toxicology” in *Occupational Medicine: Principles and Practical Applications*, 3rd ed., edited by Zeng C, Nickerson B, and Hovath EP, pp. 70–84, Copyright © Mosby/Elsevier (1994).

biotic metabolite from Phase I usually becomes inactivated. This process frequently involves “conjugation,” or the attachment of a polar functional group that makes the molecule much more hydrophilic. These functional groups include glucose (as glucuronic acid), sulfate, mercapturic acids (with a sulfur bridge), glutathione, methyl groups, acetyl groups, and amino acids, mostly glycine and cysteine.

There are numerous enzyme systems of importance in occupational and environmental toxicology. Usually, these systems are regarded as the intellectual province of research toxicologists and are treated as “black boxes” by the clinician. However, an understanding of the basic characteristics of the most important system helps one to conceptualize the process and to appreciate why some toxic agents behave as they do in the body.

### ***Mixed Function Oxidase System***

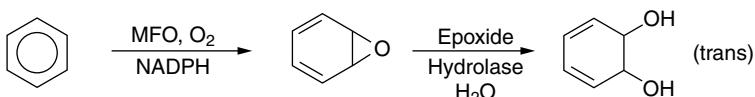
The most important of the metabolizing systems, the mixed function oxidase (MFO) system is also known by other names: aryl hydrocarbon hydroxylase, arene oxidase, epoxide hydroxylase, and cytochrome oxidase metabolizing enzymes. The MFO system is virtually ubiquitous in the body, but activity is particularly concentrated in the liver and lungs, and can be found and conveniently studied in circulating lymphocytes. The MFO system has a huge capacity and acts on a wide variety of substrates. It also has the property of being inducible; when presented with suitable substrate, the cell synthesizes more MFO enzymes, increasing the capacity of the system and preparing itself for a greater load. The degree of inducibility and the level of baseline activity in a given tissue are genetically determined; therefore, at any one time MFO activity in a particular tissue reflects heredity combined with exposure in the recent past. The MFO system is a complex of membrane-associated enzymes closely linked to the cytochrome P450 system (and other cytochromes). Various cytochromes are distributed differently in various organs. (CYP1A1, for example, is important in the oxidation of polycyclic aromatic hydrocarbons but is not present in the liver in humans. CYP2E1 is inducible by ethanol and metabolizes small hydrophobic compounds.

CYP2D6 is highly inducible, and its activity is associated with an elevated risk of lung cancer.) Figure 2.9 illustrates some of the major oxidizing reactions of the MFO system.

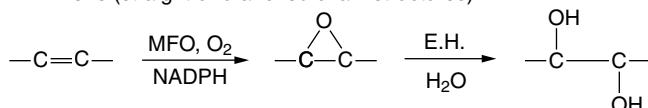
The cytochrome-linked MFO enzymes act on organic compounds with aromatic or double bonds. The system attacks these

1. Mixed function oxidase (epoxide hydrolase)

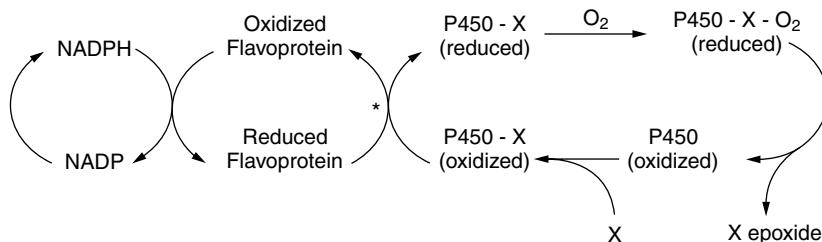
1.1 Arene (ring structures)



1.2 Alkene (straight or branched chain structures)

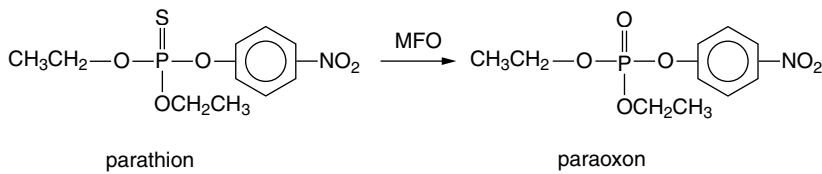


1.3 Detail of interaction between cytochrome P450 oxidase and MFO:



\* Reaction catalyzed by cytochrome P450 reductase

1.4 Transulfuration



**Figure 2.9.** Oxidizing reactions of the mixed function oxidase system. Adapted from Guidotti TL, "Principles of Occupational Toxicology" in *Occupational Medicine: Principles and Practical Applications*, 3rd ed., edited by Zeng C, Nickerson B, and Hovath EP, pp. 70–84, Copyright © Mosby/Elsevier (1994).

bonds, usually creating first an epoxide and then an alcohol, in the process first activating the compound, then deactivating it and rendering it more easily excreted.

A very common and important Phase I reaction is the formation of an epoxide out of a double-bond or in an aromatic ring structure. The epoxide intermediate is highly “electrophilic”—in other words, it is highly reactive with electron-dense “nucleophilic” compounds containing nitrogen, sulfur, carbon, and oxygen, forming a covalent bond. Because DNA is just such a compound, it is a target, or “critical macromolecule,” along with RNA and proteins. The difference is that changes in DNA may lead to carcinogenesis in a certain frequency of DNA-epoxide interactions, whereas interactions with RNA or proteins merely result in, at most, denaturation and loss of activity. The epoxide metabolite, although highly active, may be short-lived. However, its “signature” remains after Phase II, in the form of a characteristic trans-diol, a double-alcohol with the hydroxyl group on the opposite sides of the carbon-carbon bond. This occurs as a consequence of the geometry of the attack of the epoxide hydrolase enzyme, which adds a water molecule and breaks the epoxide down into two alcohols. This change in structure is called the “NIH shift,” for the institution where it was discovered.

A compound may undergo recurrent steps of metabolism. Benzo(a)pyrene, for example, may be metabolized first to a 1,2-diol and subsequently to a 1,2,3,8-tetrol (four hydroxyl groups) through two Phase I activations. The product of the second Phase I reaction (a diol epoxide) is a stronger mutagen than the first epoxide metabolite.

The interaction between the MFO system and the cytochrome P<sub>450</sub> system is intricate. The cytochromes (P<sub>450</sub> is a designation based on the spectral characteristics of the purified enzyme; there are others but none as important, and there are at least six kinds of P<sub>450</sub>) are electron-carrying links in a chain yielding energy through oxygen-dependent electron transport. This chain includes the flavoproteins, NADP, and oxygen, the final electron recipient. The MFO system acts by forming a complex between the xenobiotic and the iron-containing cytochrome P<sub>450</sub>, reducing the complex (by donating electrons to it from a flavoprotein) and then combining it with molecular oxygen,

inserting the oxygen into the bond, and stripping the xenobiotic epoxide off the P<sub>450</sub> complex, which is then regenerated. The process is fast, efficient, and remarkably nonspecific. The MFO system is capable of producing several basic organic reactions: hydroxylation of aromatic compounds, oxidation of aliphatic side chains, dealkylation of a nitrogen deamination, transsulfuration, reduction, and conjugation. Transsulfuration is particularly important in the activation of organophosphate pesticides, such as parathion, in insects as well as in mammals.

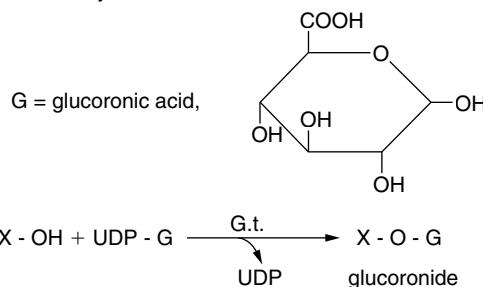
Inducibility is an important characteristic of the MFO system. There are at least two modes of induction. The best-known mode results from repetitive exposure to compounds such as phenobarbital and results in a sustained increase in activity of many MFO enzymes and an increased content of cytochrome P<sub>450</sub>. The other recognized mode responds to single exposures to compounds like 3-methylcholanthrene, benzo(a)pyrene, and 2,3,7,8-tetrachlorodibenzodioxin in a rapid but selective increase in activity of a few enzymes and other cytochromes, with designations like P<sub>1450</sub> and P448.

### *Glucuronyl Transferase*

This enzyme system is a mechanism for attaching a glucose molecule onto a metabolite or xenobiotic with an exposed “nucleophilic” (electron-dense) site, usually a hydroxyl (-OH) but sometimes an amine (-NH), mercaptan (-SH), or carbonyl (-CH). The result is a glucuronide derivative that loses its toxic activity. This conjugation also renders the molecule very water soluble and easily excreted into urine or bile. The enzyme system is concentrated in the liver and is also found in the kidney, skin, and gastrointestinal tract, but not in the lung. Figure 2.10 illustrates the action of this enzyme system.

Glucuronyl transferase acts on many xenobiotics that have already been through Phase I metabolism and have been converted into alcohols. It accepts a very wide range of such compounds and is unusually nonspecific in its substrates for a metabolizing enzyme, but it is very specific for the ligand UDP-glucuronic acid (UDP is a carrier protein complexed to the acid derivative of glucose) and will not

## 2. Glucuronyl transferase



## 3. Sulfotransferase



PAP-S = 3'-phosphoadenosine - 5'-phosphosulfate  
ABP = adenosine - 3', 5'-biphosphate

**Figure 2.10.** Conjugation reactions in Phase II biotransformation.

accept any other sugar or sugar derivative. The enzyme is buried deep in the lipid bilayer of the plasma and nuclear membrane, and its activity is lost if its lipid matrix is disrupted. However, UDP-glucuronic acid is very hydrophilic, and the way the complex manages to get from the cytoplasm deep into the membrane, where the reaction occurs, is a bit of a mystery. The system acts primarily on lipophilic xenobiotics.

An important exception to the general rule that the system and its products are a mechanism of detoxification, not activation, is the conjugation of *N*-hydroxyarylamines, Phase I metabolites of arylamines, including aniline dyes and  $\alpha$ - and  $\beta$ -naphthylamine. After conjugation, the glucuronide is very hydrophilic and is excreted in the urine. In the bladder, the glucuronide is split off by an enzyme ( $\beta$ -glucuronidase) active at low pH (5), which leaves an aryl nitrenium (a ring structure with an exposed, positively charged  $-\text{N}=\text{O}^+$  group). This aryl nitrenium is intensely electrophilic and carcinogenic—leading to the induction of bladder cancer—but it would not have been produced without the prior action of glucuronyl transferase.

### ***Sulfotransferase***

This is the other major conjugating system, linking the xenobiotic or its metabolite to sulfate. The system consists of many highly specific enzymes catalyzing similar reactions in different substrates. They act primarily on more-hydrophilic substrates and are present in cytoplasm, not in the membrane. The reaction products are highly water soluble and easily excreted. The system uses PAPS, a key sulfur donor in the cell that consists of a sulfate linked to adenosine phosphate. PAPS is regenerated from ATP and sulfate, which is abundant in the cell. Sulfotransferase activity is proportional to the amount of sulfur taken in the diet and may have evolved as a way of handling excess sulfur. Figure 2.10 illustrates the sulfotransferase reaction.

### ***Glutathione Transferase***

This system, known confusingly by several other names (including glutathione alkyl/aryl transferase, steroid isomerase, ligandase, and others), consists of an enzyme acting on a very wide range of substrates with only their electrophilic nature in common, although lipophilic compounds are favored. Glutathione is a tripeptide ( $\beta$ -glutamyl-cysteinyl-glycine) with an exposed sulphydryl (mercaptan) group (on the cysteine). The glutathione is bound to the xenobiotic as a conjugation reaction on the sulphydryl group; the peptide is then cleaved, leaving a cysteine residue attached to the metabolite, a mercapturic acid or (if the substrate was aliphatic) a thioether. The enzyme system is most heavily concentrated in the kidney but is also present in the liver and in small levels in other tissues. Massive quantities of the enzyme may be present in the cytoplasm of the cell. It can be induced and is also under hormonal control.

Oxidant gases are encountered in situations such as welding fumes, chemical leaks, uncontrolled plastics fires, sparking electronic apparatus, and hyperbaric oxygen. Oxygen itself has an oxidizing effect on cells. Exposure to oxidant gases, such as nitrogen dioxide, perturbs the antioxidant-reduction system of the cell, causing depletion of glutathione reserves (a useful marker) and antioxidant reserves (particularly

ascorbate and tocopherol). The susceptibility of a given individual to the various end points possibly varies following exposure to such gases. Determinants of the variation of these end points may depend on the antioxidant reserve in the individual and on the distribution of such reserves in the population.

Another glutathione-dependent system is thioltransferase. This ubiquitous NADP-dependent enzyme system regenerates thiols (sulfhydryl groups, -SH) that have been converted under conditions of oxidant stress or depletion of intracellular glutathione to disulphide linkages (-SS-). Disulphide bonding of adjacent or nearby sulfhydryls is an important mechanism of physiologic control of some enzymes, but when it occurs during excessive oxidation it can result in deactivation of enzymes and disulphide poisoning. The system is thought to have evolved as a means of adaptation to aerobic metabolism by primitive bacteria.

### ***Glutathione Peroxidase***

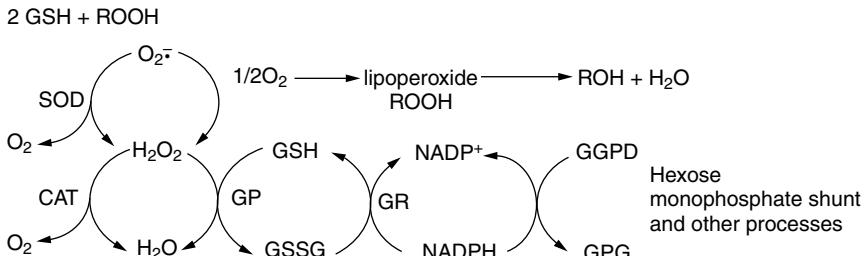
Yet another system based on glutathione, glutathione peroxidase is an extraordinarily fast-acting selenium-containing metalloenzyme that was originally confused with glutathione transferase. It acts in concert with superoxide dismutase and catalase to reduce peroxide, as shown in Figure 2.11. It is ubiquitous in mammals; it has a particular physiological role in maintaining the integrity of the erythrocyte cell membrane against oxidation; and its activity varies in each organ and is influenced by numerous factors, including dietary selenium, tocopherol intake (vitamin E), age, sex, oxidant exposure, presence of peroxidized lipids in the diet, and hereditary factors. It can be found in the cytoplasm, with some activity in mitochondria.

### ***Other Antioxidant Systems***

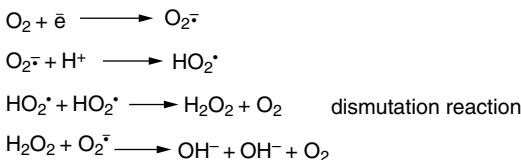
Other antioxidant systems are also included in Figure 2.11. Catalase is an abundant enzyme that is an integral part of the body's protection against oxidants and free radicals, reducing hydrogen peroxide to

CAT = catalase  
 SOD = superoxide dismutase  
 GP = glutathione peroxidase  
 GR = glutathione reductase  
 GS<sup>•</sup> = glutathione

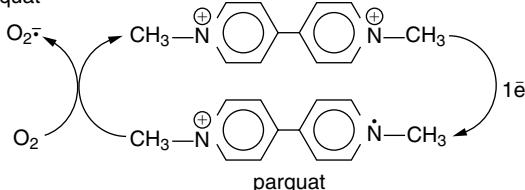
O<sub>2</sub><sup>•</sup> = superoxide anion  
 OH = hydroxyl radical  
 O<sup>•-</sup> = singlet oxygen



#### Generation of Superoxides



#### Paraquat



**Figure 2.11.** Superoxide formation and reduction. Adapted from Guidotti TL, "Principles of Occupational Toxicology" in *Occupational Medicine: Principles and Practical Applications*, 3rd ed., edited by Zeng C, Nickerson B, and Hovath EP, pp. 70–84, Copyright © Mosby/Elsevier (1994).

water and molecular oxygen. It is also capable of converting ethanol to acetic acid in the presence of H<sub>2</sub>O<sub>2</sub>, which is relatively stable.

A widespread family of at least two enzymes particularly rich in the lung, superoxide dismutase (SOD) is another critical component of the body's antioxidant mechanism. It acts quickly and depends on a metal ligand (copper or manganese) to reduce the superoxide radical, converting it to an oxygen molecule and a hydrogen peroxide mole-

cule in the presence of water; the hydrogen peroxide is later reduced by catalase or glutathione peroxidase, which also acts on lipoperoxides. Because the superoxide and hydroxyl radicals attack lipids and macromolecules avidly, they produce great damage to membranes and DNA. Each is highly unstable. The system probably developed as an adaptation to an aerobic environment in primitive life. Toxicologically, it protects against the generation of free radicals by ionizing radiation or by free radical-forming agents, such as paraquat.

### *Alcohol-Metabolizing Systems*

Alcohol dehydrogenase is the principal pathway for metabolizing alcohols, particularly ethanol. This cytosolic enzyme is found in abundance in the liver. It is highly unusual in that it is the only important example of an enzyme with apparently zero-order kinetics: between 100 and 200 mg/dl of blood, ethanol is metabolized at a relatively constant rate regardless of the plasma concentration. This phenomenon is described in detail under the heading "Kinetics." Ethanol is metabolized to acetaldehyde, which is further metabolized in the mitochondria by the action of aldehyde dehydrogenase to acetic acid, which is then disposed of in the tricarboxylic acid cycle. Aldehyde dehydrogenase is inhibited by disulfiram, a constituent of treated rubber, and the accumulation of acetaldehyde results in unpleasant but harmless symptoms. This was exploited as a preventive measure to discourage alcohol abuse, in the form of the drug Antabuse. When the substrate of alcohol dehydrogenase is methanol, the product is formaldehyde, which is then converted to formic acid, causing blindness due to retinal toxicity and metabolic acidosis.

The microsomal ethanol oxidizing systems (MEOS) constitute an alternative pathway for the metabolism of alcohols to acetaldehyde, especially ethanol. It appears to be significant for bulk removal of alcohol at high concentrations—in the life-threatening range—and may be important in the metabolism of other compounds physiologically, although its role is only incompletely understood. Like alcohol dehydrogenase, MEOS present a picture of zero-order kinetics, but in reality this is only an approximation—the kinetics of both

are actually first-order, but the response is rather flat and the system is quickly saturated. The implications of zero-order kinetics are described below.

### ***Reducing Systems***

Numerous systems exist for the reduction of xenobiotics and their metabolites, the reversal of the direction of the enzyme systems described above. Too numerous to be discussed individually, they include esterases, amidases, glycosidases, and other hydrolytic enzymes, as well as reductases acting on aldehydes, ketones, and azo- and nitro- compounds.

One important member of this group is a system-catalyzing reduction and dehalogenation in organohalides, particularly chloro- and bromocarbons, fluorocarbon bonds being too strong to cleave. This system, for example, converts DDT to DDD (lacking one chlorine at the central carbon).

Hydrolytic reactions are very important but nonspecific. It is thought that virtually any tissue can hydrolyze virtually any hydrolyzable substrate. The toxicological role of these enzymes is a byproduct of their physiological role in intermediary metabolism. Esterases are particularly important toxicologically because this enzyme class regenerates the acetylcholinesterase enzyme at the synaptic level when the active site is occupied by an organophosphate insecticide, such as parathion.

### ***Example: The Biotransformation of Benzene***

Benzene is a convenient example of biotransformation, both because it is the simplest aromatic molecule and because its pathways illustrate many common features and reactions. It also has important implications because benzene, once activated, is an important toxic agent, causing aplastic anemia in bone marrow and acute myelogenous leukemia. Figure 2.12 presents the metabolic pathways of the biotransformation of benzene.

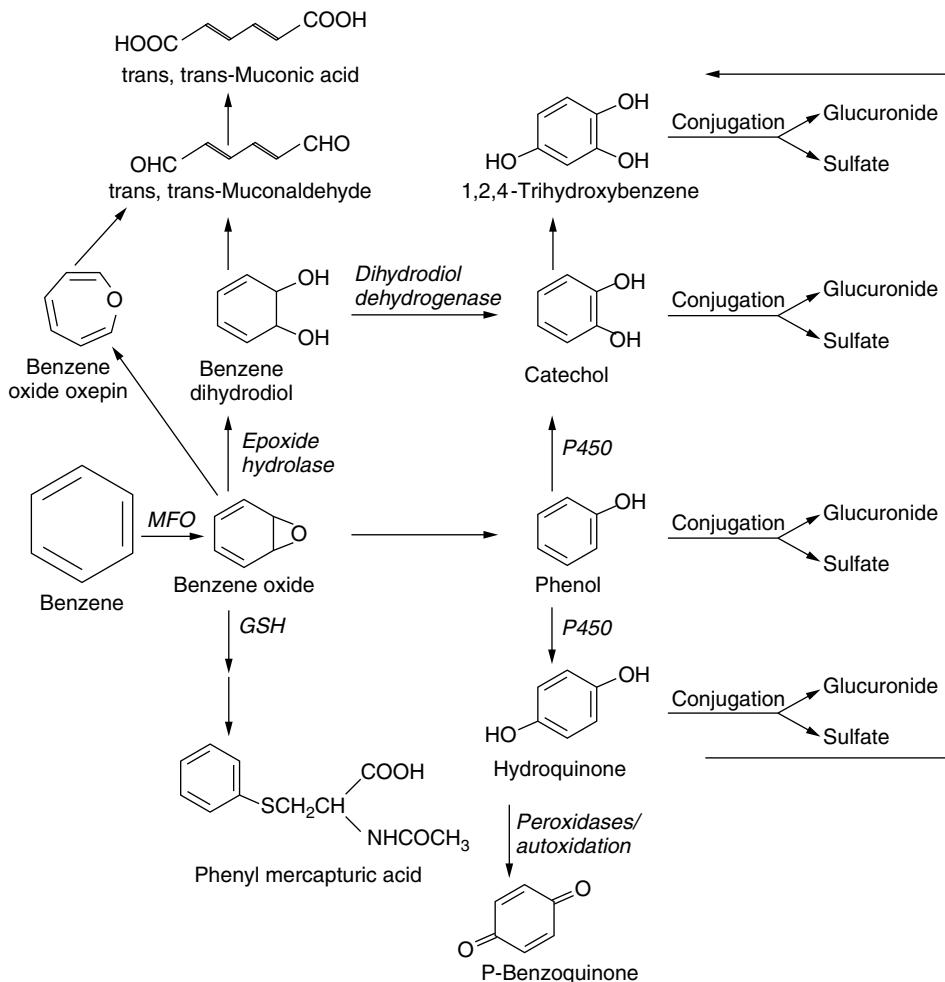


Figure 2.12. Biotransformation of benzene.

Like many aromatic compounds, the initial step in the metabolism of benzene is the formation of an epoxide, which may go in four directions. The epoxide may be sulfated into a mercapturic acid, which is soluble and easily excreted. The epoxide may pass into an unstable form in which the oxygen is incorporated into the ring and then open into a chain, called muconic acid. The epoxide may also hydrate and

convert to a diol (benzene glycol), producing the classic “NIH shift,” which may also result in opening the chain. The epoxide may also form an alcohol, producing phenol. Phenol and the diol (benzene glycol) may form catechol, which goes in one direction, or phenol may form a quinone, which goes in the other direction. The quinones, catechols, catechol’s derivative trihydroxybenzene, and phenol itself may all be conjugated as either the glucuronide or the sulfate.

Thus, the biotransformation of benzene can proceed in many directions. In practice, however, they are not equally important. The pathways that lead to benzoquinone and to muconic aldehyde are the main ones. The mercapturic acid pathway, which leads to glutathione conjugations, is minor. All the conjugates—muconic acid, phenol, and benzene—appear in the urine.

For purposes of biomonitoring, urinary benzene and trans, trans-muconic acid are preferred because there is a direct relationship with exposure. The measurement of benzene in urine is a good indication of benzene exposure during the previous 24 hours.

For the OEM physician who is not trained as a toxicologist, a little time spent studying the biotransformation of benzene is an investment in learning patterns of biotransformation that apply to many organic compounds.

## **Excretion**

The xenobiotic or its metabolites would accumulate and remain within the body if there were no mechanisms for excretion. Excretion mostly occurs through the kidney, the liver and biliary system, the gut, and the lungs. Minor pathways include lactation, transfer to the fetus through the placenta, and secretion through specialized exocrine glands (sweat and saliva).

The kidney is the major route of excretion for most xenobiotics. Those that are water soluble may be filtered or excreted unchanged. The reserve capacity of the kidney is very great, and this mechanism is rarely saturated in healthy people. But individuals with renal insufficiency may show accumulation and persistence of the xenobiotic

and, consequently, prolonged and more severe toxicity. Other xenobiotics may be metabolically transformed into more water-soluble metabolites before renal clearance occurs. Xenobiotics that are themselves nephrotoxic may injure the kidney and reduce their own clearance, enhancing their own toxicity by further accumulation.

The liver, besides being an important metabolizing organ, secretes some xenobiotics into bile, including metals and some organochlorine compounds. They are then carried into the gut and excreted with feces. Some of these products depend on bile production for excretion, but others are independent of bile flow. Mercury and thallium enter the bile through simple diffusion. Conjugated organic compounds, arsenic, and lead are actively secreted. Chromium, iron, zinc, and proteins are actively excluded from bile, presumably because they resemble molecules the body has evolved mechanisms to retain. The presence of bile emulsifies lipophilic substances but also binds and transports polar compounds and metals. Because bile is difficult to study—being very inaccessible in the abdomen—hepatobiliary excretion is not as well studied as other routes. It is known, however, that thresholds for excretion are different between human beings (500–700 daltons) and rats (325 daltons).

Some xenobiotics and their products are excreted through the bile but are then reabsorbed, either directly or after they pass into the gut and are deconjugated by the metabolism of gut flora or hydrolysis in the intestine. This is called the “enterohepatic circulation.” Once absorbed into the circulation, xenobiotics may return to the kidney to repeat the cycle of excretion and reabsorption. These compounds may recirculate by enterohepatic circulation many times, persisting in the body much longer than they otherwise would, or they may pass out of the body in feces. Forced biliary excretion is not presently possible, but interruption of the enterohepatic circulation by binding agents such as cholestyramine is a practical clinical intervention used to hasten excretion and reduce the body burden of xenobiotics excreted in the bile and reabsorbed in the gut. This intervention was first demonstrated for kepone, a discontinued pesticide. Although hepatotoxic agents may interfere with their own excretion by the

liver, they are more likely to interfere with metabolism, and as a practical matter this effect is rarely significant.

Intestinal excretion is relatively slow. Some luminal cells are sloughed, taking with them some substances, such as iron. However, there is also some evidence that the gut may excrete or allow diffusion of xenobiotics directly into the lumen and out via the feces. The process is probably affected by gut flora and diet.

Volatile gases are readily excreted by the lung through passive diffusion from the blood, crossing the alveolar-capillary barrier in “reverse” direction. Gases that are poorly soluble in blood, such as ethylene, are rapidly and efficiently eliminated by this route, including many solvents. Those that are readily soluble in blood, such as chloroform, are less efficiently eliminated and may be detectable in expired air for days or even weeks. Respiratory excretion may also be affected by the affinity of the gas for hemoglobin (in the case of carbon monoxide) and by minute ventilation. Some compounds, such as organic ethers, are principally or almost entirely eliminated by the lung. Others are eliminated to a negligible degree toxicologically but enough to permit traces of the gas to be detected in expired air. Because the volatile gas diffuses directly out of the blood into the alveolar space, concentrations of the gas at the alveolar level accurately reflect the concentration of the dissolved gas in blood. This has led to a new and very practical, noninvasive approach to quantifying exposure levels by recovering expired air (in particular end-expired air, which must closely approximate alveolar concentration of the gas) and measuring the concentration of the agent by mass spectroscopy. The technique is very sensitive, but detection levels, duration of detection, and elimination curves over time are highly specific for each gas. Ethanol is also excreted in this manner, and expired alveolar air closely reflects blood concentration, making the familiar breath monitors for alcohol a very effective biomonitoring instrument.

Mechanisms for excretion of particles are described later in the subsection on inhalation toxicology.

Xenobiotics and their metabolites are also eliminated by various minor routes that matter little with respect to reduction of the total

body burden, but that may have some toxicological implications. Lipid-soluble agents may be secreted in breast milk. Lactation is both a major route of exposure of infants and young children to substances such as the organohalides, including PCBs, and a route of excretion for the lactating woman for the same compounds. Water-soluble agents are excreted in saliva and tears and are filtered through sweat glands, the latter functioning much like individual nephrons. Lipid-soluble agents may also be found in cerumen and sebum. These minor elimination pathways permit noninvasive monitoring techniques for the detection of the agent but, aside from saliva, are not present in sufficient volume to quantify exposure. Salivary levels are usually at about 10 percent of the level of similar xenobiotics in plasma. Minor pathways of excretion, particularly saliva, are likely to play a greater role in biomonitoring in the future.

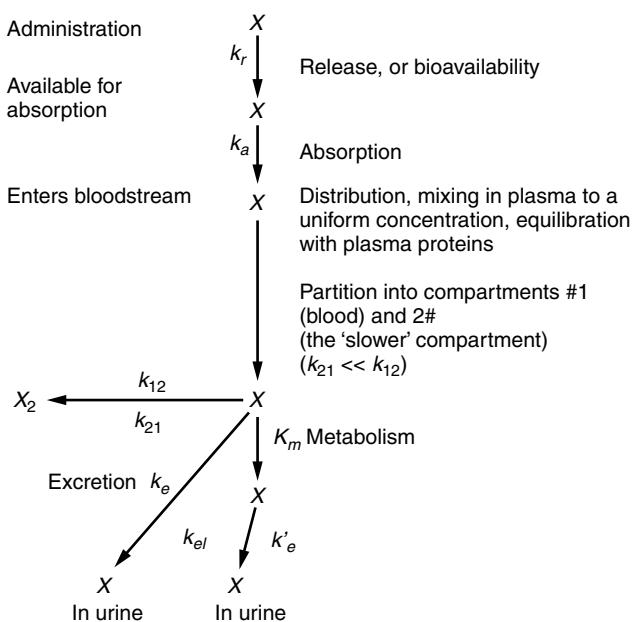
## Kinetics

“Elimination” is the process of removing the xenobiotic or its metabolites from blood and tissue, which depends on blood for the mobilization of stored xenobiotics and delivery to organs of excretion. Metabolism and excretion, singly or together depending on the xenobiotic, determine the rate of elimination from the blood, which can be monitored and used to infer the pattern of elimination, which is an important clue to how the xenobiotic is behaving and where it is going in the body. One of the most important factors in the kinetics of a xenobiotic is how it partitions into various “compartments” of the body besides plasma, such as adipose tissue, bone, liver, and so forth. A compartment may be thought of as an anatomic or functional space (it may not be “real” or tangible; for example, bound xenobiotic circulating in plasma). There are two essential flow rates ( $Q$ -dot), determined by the movement of mass and volume over time—one into and one out of the compartment—as well as flux (usually not rate limiting) rates between the interstitium of tissue and the intracellular compartments. The individual compartments may be characterized by volume, density, weight, fat content, and many physiological

parameters such as renal clearance rates (in the kidney) and alveolar ventilation (in the lung). The more compartments, the more complex the behavior and, therefore, the more complicated the toxicokinetic model required to describe its behavior in the body, as discussed later.

Toxicokinetics is most easily understood using the example of a single bolus of xenobiotic administered by intravenous injection, although this example is more relevant to pharmacology than occupational and environmental toxicology. Figure 2.13 describes each step in the absorption, distribution, metabolism, and excretion of a xenobiotic in terms of rates specific to each compartment but superimposed in the elimination curve in plasma.

Returning to Figure 2.2, the concentration in the circulation over time, the “area under the curve” (AUC) is constant and reflects the portion of the total dose delivered that is present in plasma or the compartment being studied. The AUC is the integral over time and



$k_{el}$  = elimination rate, is a summary rate for elimination of  $X$  by all pathways

**Figure 2.13.** Rates and toxicokinetics.

is equal to the total mass, or the dose, both taken in and eliminated by all routes.

Each compartment's rate of elimination is described by a rate constant ( $k$ ) that determines how rapidly the process proceeds. The rates of eliminations are expressed in units of inverse time, as in "per second." The mathematics of clearance, describing the concentration in plasma  $C$ , are based on a simple decay curve that describes turnover in the compartment, but the mathematics soon become more complicated as more compartments are added, as follows:

$$\begin{aligned} C_0 &= \text{Dose}_{\text{iv}} / V_d \text{ at time } 0 \\ C_t &= C_0 \times e^{-kt} \text{ at time } t, \text{ where} \\ &\quad \text{the elimination rate } k \text{ has units of } t^{-1} \end{aligned}$$

This equation describes a "logarithmic decay" as the concentration in plasma falls over time. The complexity then occurs when there are additional compartments ( $1, 2, 3, \dots, n$ ) and the xenobiotic is distributed among them proportionately ( $F$ ). Each compartment has its own rate ( $k$ ) reflecting elimination from the compartment and communication with the plasma compartment. The description of the final concentration in plasma at time  $t$  would then be:

$$C_t = F_1 C_0 e^{-k_1 t} + F_2 C_0 e^{-k_2 t} + F_3 C_0 e^{-k_3 t} + \dots + F_n C_0 e^{-k_n t}$$

This equation is fundamentally the sum of clearances and releases back into plasma of all the compartments. Graphically, the curve of such an equation resembles that in Figure 2.E, taking time = 0 to be the peak. It describes a complicated curve in which absorption and distribution play their respective roles in achieving the peak concentration, and redistribution and elimination (both excretion and biotransformation) follow the peak. Elimination curves plotting blood concentration against time define the single, sequential, or overlapping rates of change of the concentration of the xenobiotic in blood over time, and each rate is described by a rate constant ( $k$ ) that determines how rapidly the process proceeds. Each phase of the

kinetics of a xenobiotic is determined by properties of the agent and characteristics of the biological system.

“Clearance” describes elimination in terms of the equivalent volume of plasma from which the xenobiotic or its metabolite would have been completely removed, as the reduction in concentration proceeds. The volume of distribution,  $V_d$ , as in pharmacology, is an idealized, calculated volume in which the xenobiotic distributes itself, not an actual space. If a xenobiotic confined itself to the vascular space, the  $V_d$  would be the total volume of plasma. If a xenobiotic leaves the vascular space and is found in extracellular fluid or is accumulated in a storage depot, the  $V_d$  will be larger than the plasma volume. Formally, clearance is equivalent to the total dose divided by the AUC (area under the curve) for the various compartments ( $1, 2, 3, \dots, n$ ) or the product of the volume of distribution and the elimination rate:

$$Cl_{1, 2, 3, \dots, n} = \text{Dose}_{\text{iv}} / \text{AUC}_{1, 2, 3, \dots, n} = V_d k,$$

for a given compartment;

$$Cl = Cl_1 + Cl_2 + Cl_3 + \dots + Cl_n,$$

for all compartments.

The single most important implication of the clearance equation, however, is that it allows the calculation of a critically important and much-used summary indicator of clearance, the “half-life,” which is the length of time it takes for the concentration of a xenobiotic to fall by one half. For the simple case:

$$t_{1/2} = 0.693 \times V_d / Cl$$

Rate constants are also described by their “order” (essentially, the exponent of the rate) and complexity, as indicated by the number of defined compartments. When the elimination pathway is working at less than capacity, the kinetics is usually first order. When the system is saturated and the capacity is exceeded, the kinetics is zero order. When there is more than one compartment with different kinetics,

the kinetics is second order or higher and the model is referred to as “multicompartment.”

A zero-order rate constant describes an elimination curve in which the rate is limited intrinsically by a fixed capacity of the body to eliminate the agent, regardless of its concentration. In practice, the only important example of this is, ironically, the most common practical problem of toxicological concern: alcohol dehydrogenase (described above in the subsection on metabolism), which metabolizes ethanol, methanol, and other alcohols. Regardless of how much alcohol a person takes in, elimination will occur at the same rate. This is because the metabolizing enzyme system is quickly saturated, and thereafter there is a bottleneck in the reaction: only so much mass can be biotransformed over time. Zero-order kinetics is a sign of saturated capacity. The rate is fixed and constant over time and across most of the range of concentrations of the xenobiotic. There is no true half-life because the concentration keeps dropping at a constant rate. For ethanol, that rate is generally 20 to 30 mg/dl/hour, which allows easy back calculation to concentrations earlier in the day, a great convenience in forensics.

A first-order rate constant describes a process in which the rate of elimination is proportional to the concentration of the agent in plasma—the most common situation, especially for water-soluble xenobiotics. This is also called a “one-compartment” model because the agent behaves as if it is restricted to one compartment of the body, usually the vascular space. A one-compartment model assumes that there is rapid equilibrium in the plasma (no mixing time) and that distribution in the body is more or less homogeneous, although in fact there are different concentrations of the xenobiotic in different tissues. This is the normal situation: the elimination pathway operates below capacity, and the mechanism of excretion or biotransformation is driven, thermodynamically, by the concentration of the xenobiotic, because that determines how frequently a molecule will come into contact with the mechanism and therefore will be available to be acted upon. Many xenobiotics, such as 1,3-butadiene, follow first-order kinetics. The rate is not fixed but instead varies by

the concentration of the xenobiotic in the compartment and is described by an elimination rate constant ( $k$ ) that is the exponent in the clearance equations. The parameters  $k$ ,  $V_d$ , and the resulting  $t_{1/2}$  are independent of concentration or dose. The curve is a simple logarithmic decay. The period required for the plasma concentration to drop by half is described by a single half-life ( $t_{1/2}$ ). The  $t_{1/2}$  can be calculated easily and accurately and is related to the elimination rate by  $t_{1/2} = 0.693 / k_e$ . Many occupational exposures follow this pattern.

Multicompartment or higher-order toxicokinetics suggests that the agent equilibrates in more than one compartment and is eliminated at different rates from each. The elimination will not fit a simple logarithmic decay (or straight line on a logarithmic scale) but must be described by a more complex equation described by at least two rate constants—a “fast” rate constant and a “slow” one, each with its own half life. There may be several pathways with several rate constants and several half-lives, each of which apply to a different segment of the elimination curve. Among organic compounds, styrene illustrates this behavior very well. There is more than one rate of elimination, which is expressed as a change in the shape of the curve over time. Organohalides typically show second-order kinetics, and because of their storage and slow release from fatty tissue, they will show a very long “tail” to the curve, in which plasma levels may persist for a very long time. This is the most common situation in environmental toxicology.

Multicompartment models based on physiology are used to predict the toxicokinetics of all but the simplest xenobiotics. The most advanced of the models are called “physiologically based pharmacokinetic models” (PBPK), with no distinction made for their application to toxicokinetics. In addition to their value in research and in pharmacology, these models are important in risk assessment, making extrapolations from animal species to humans, calculating what exposures might have been in the past (for persistent xenobiotics), and adjusting complex exposure profiles in setting standards or adjusting them to, for example, different work-shift patterns. These models can be excruciatingly complicated, but they are built on

simple elements. For each compartment, there is a flow in and flow out, perfusion rate, affinity for the xenobiotic, and metabolism rates, and for the organs of excretion, there are clearance rates. These are measured or estimated and strung together into comprehensive models. These models are computationally complicated more because of redundancy, physiological complexity, and the huge number of parameters than because of mathematical complexity.

Figure 2.13 presents a conceptual framework for understanding kinetics of elimination.

## TOXICODYNAMICS

Toxicodynamics refers to the interaction of the xenobiotic with the body and the body's response. This section will introduce basic principles of exposure-response that are of profound significance to occupational and environmental medicine. The emphasis will be on concepts that are clinically significant and that are important in interpreting the risk profile of a particular chemical hazard.

Individual mechanisms of toxicity are too varied to discuss in detail in this chapter. Rather, some will be mentioned in passing as examples and others will be discussed under the entries for particular xenobiotics and classes of chemicals in Chapter 10. Inhalation toxicology is sufficiently specialized, and essential to OEM, to merit its own section. Skin toxicity is more easily discussed in the overall context of occupational disease in Chapter 17.

### General Mechanisms of Toxicity

Toxicodynamics can be thought of as a stepwise process characterized by phases, in the same way as toxicokinetics (a formulation given in the popular textbook by Casarett and Doull). The first phase might be delivery of the xenobiotic to tissue (embracing the phases in toxicokinetics of absorption, delivery, and possibly metabolism, when a xenobiotic is biotransformed into a more active “ultimate toxicant”). The second phase would be interaction of the active agent

with a target molecule. The third would be the consequences of cell dysfunction and the expression of toxicity, which, once they occur, lead to secondary changes, responses, and common pathways of pathology. The fourth phase would be restoration of function and repair, which may or may not be adequate or functional.

Xenobiotics exert their effects by so many different mechanisms of action that it can be said that toxicology studies, in essence, the interference of any natural function in the body by chemicals and response. Common mechanisms of toxicity include interaction with specific receptor molecules, inhibition of enzyme function, cell dysfunction, lipid peroxidation, induction of cell death (apoptosis), immune dysfunction, and genotoxicity. There are also common pathways that result from specific mechanisms of exposure, among them inflammation.

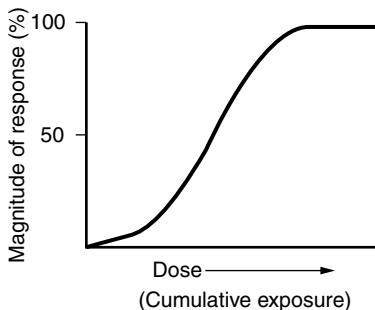
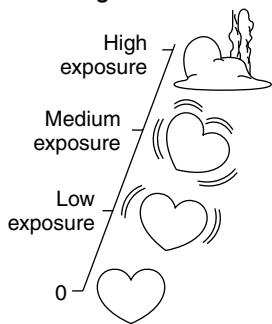
### **Three Exposure-Response Relationships**

The exposure-response relationship is a concept fundamental to an understanding of toxicology. It was Paracelsus himself who first said, “It is the dose that makes the poison,” and thereby established that poisons were not a mystically benighted form of matter but that all chemicals had toxic properties that become apparent as increasing quantities are consumed or absorbed. It follows from this simple observation that there may be “safe” levels of exposure to even the most toxic substances, a much more controversial assertion. Obviously, there are several dimensions to this seemingly straightforward concept. There are three distinct varieties of the exposure-response relationship that need to be distinguished conceptually. These are the toxicological dose-response relationship, the clinical dose- or exposure-response relationship, and the epidemiological exposure-response relationship. These are illustrated in Figure 2.14.

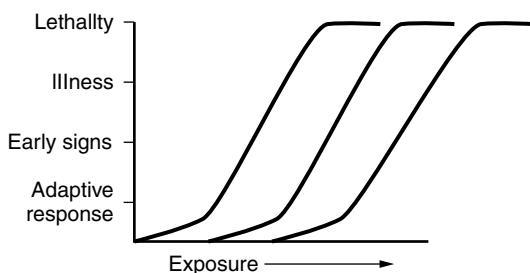
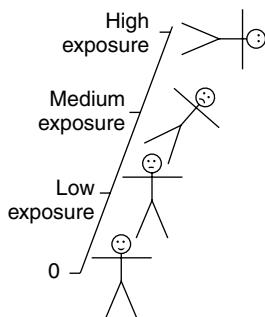
“Dose” is generally understood to mean the total mass of a xenobiotic administered to the subject, such as a pill taken by a person or a quantity injected into an animal in a laboratory experiment. Dose also refers to the cumulative amount of the xenobiotic delivered to

## EXPOSURE-RESPONSE RELATIONSHIPS

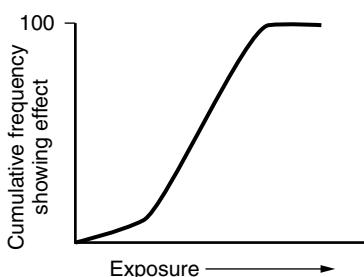
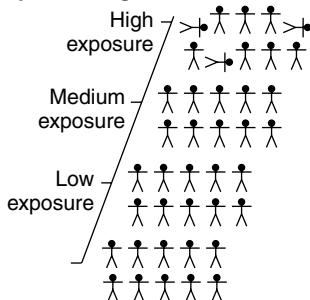
### Toxicological



### Clinical



### Epidemiological



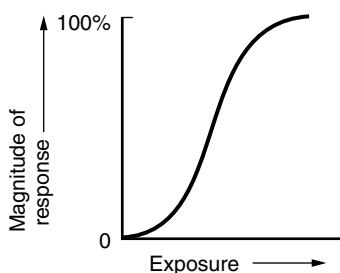
**Figure 2.14.** Three exposure-response relationships: toxicological, epidemiological, and clinical. Adapted from Guidotti TL, "Principles of Occupational Toxicology" in *Occupational Medicine: Principles and Practical Applications*, 3rd ed., edited by Zeng C, Nickerson B, and Hovath EP, pp. 70–84, Copyright © Mosby/Elsevier (1994).

the subject over the period in which it was given. “Exposure” is generally considered to be the level of concentration in a medium in which the xenobiotic is available for absorption, by any given route or all routes together, at or over a given period of time. In OEM, unlike medicine in general, exposure is almost always of greater interest than a single dose, and one is most concerned with the concentration, which usually varies over time. In this context, dose is of interest only as the cumulative amount that is absorbed after variable exposure—the internal dose taken up by the body.

Dose, in the context of OEM, is cumulative exposure. If the dose is given all at once, the term “dose-response relationship” is most meaningful, as it is when the toxic substance is accumulated in the body. If the exposure takes place over a prolonged period of time, the internal dose at any given time tends to vary, and it is more useful to think of an “exposure-response” relationship.

A particularly important, if confusing, term in toxicology is “threshold,” which means the level of dose or exposure at which an effect is first observed. The existence of thresholds for certain types of response (particularly carcinogenesis) are controversial, and the arguments surrounding identification of a threshold for response frequently neglect to specify the type of threshold under consideration.

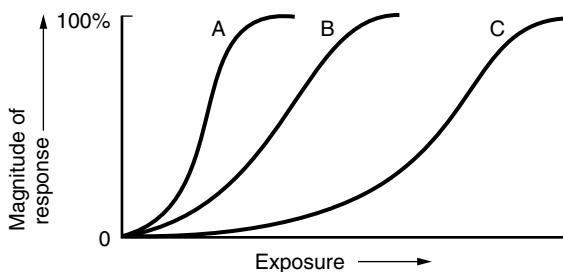
The most fundamental building block of toxicology is the dose- or exposure-response relationship demonstrable in the laboratory, which is often called the “toxicological” dose- or exposure-response relationship (see Figures 2.14 and 2.15). The relationship demonstrates



**Figure 2.15.** The toxicological exposure-response relationship. Adapted from Guidotti TL, Exposure to hazard and individual risk: When occupational medicine gets personal. *Journal of Occupational Medicine*, 1988; 30: 570–577. Reprinted with permission.

that increasing exposure or dose results in an increase in the magnitude of the effect over a broad range. The toxicological dose-response relationship, familiar from experiments in pharmacology, is similar to the experiment in which a smooth muscle agonist is added to a saline bath in which is placed a strip of guinea pig ileum, or a ring of trachea, attached to a strain gauge. The more agonist added, the more the smooth muscle contracts, at first slightly, at the point of the “threshold” (the earliest physiological or toxicological response), then in a response proportionate to dose over a broad range, and finally inefficiently near maximal contraction, generating a sigmoid curve. Similar dose-response relationships exist for other physiological responses and other toxic or pharmacological agents, often in the same organ. The fundamental principle is that the physiological response depends on the amount of the agent presented to the tissue. The dose- or exposure-response relationship has a typical sigmoid shape because the mechanisms that produce the effect are almost always less efficient at the extremes and most efficient in the middle of the range. Organophosphate pesticide toxicity, progressively inhibiting more cholinesterase at the tissue and neurotransmitter level and producing increasing cholinergic activity, demonstrates the increasing magnitude of a single effect with increasing dose or cumulative exposure.

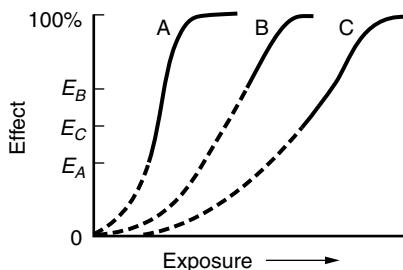
In the intact organism, or patient, exposed to a toxic substance, there are many toxicological dose- or exposure-response relationships that occur simultaneously in different organs and tissues (see Figure 2.16). Most have no clinical significance because they result in no adverse



**Figure 2.16.** Multiple toxicological exposure-response relationships arising in the same subject. Adapted from Guidotti TL, Exposure to hazard and individual risk: When occupational medicine gets personal. *Journal of Occupational Medicine*, 1988; 30: 570–577. Reprinted with permission.

effects, pathological changes, or clinical response within the range of exposure, although they may be useful as indicators for biological monitoring purposes. A few effects are critical to the health and survival of the organism. Typically the most critical or vulnerable organ system will demonstrate the first toxic effect that can be observed clinically, although in the case of skin it may be the most visible organ. As a result, the literature of clinical toxicology refers frequently to thresholds of response—doses or exposures that are associated with specific clinically apparent outcomes. These are not the same thresholds as understood in toxicology, which refer to the initiation of an effect at the cellular or physiological level. In a given individual, exposure to an increasing amount of a toxic substance leads to the progressive appearance of new and usually more severe health problems, leading at the extreme to death in a sort of stepladder to lethality. For example, the toxicity of hydrogen sulfide, which has a very steep exposure-response relationship that is driven by concentration and only secondarily by duration, progresses with increasing exposure through low-grade irritation, olfactory paralysis, acute loss of consciousness, pulmonary edema, and lethality. The toxicological exposure-response relationship, to sum up, describes the magnitude of effects in cell, tissue, or intact organism proportionate to exposure.

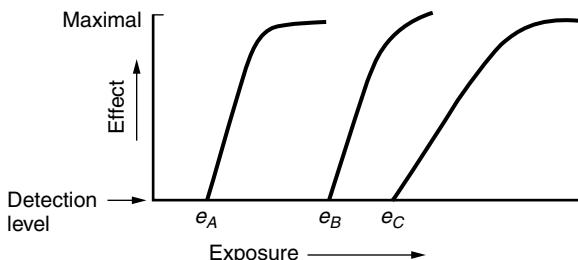
The detection of the expected pathological changes or clinical response depends on the sensitivity of pathological examination and clinical and laboratory tests. Clinical tests are often inadequate for early detection of equivocal cases because they were almost invariably designed to support making specific diagnoses in people known to be sick, with a high probability in the first place of having a particular type of disease. This contributes further confusion over the definition of a threshold. If there is a phase of subclinical adaptation to the effects of the xenobiotic, or if the onset of an effect is subtle, a clinical test may not pick it up early. For example, although nerve conduction slows with increasing intracellular lead concentration in the neuron, clinical nerve conduction studies do not demonstrate an abnormality in heavy metal toxicity until the process is moderately advanced. The early part of the exposure-response relationship is



**Figure 2.17.** Because of insensitive measurement, much of the early, low-level segment of the toxicological exposure-response relationships is hidden from view. Adapted from Guidotti TL, Exposure to hazard and individual risk: When occupational medicine gets personal. *Journal of Occupational Medicine*, 1988; 30: 570–577. Reprinted with permission.

simply hidden (see Figure 2.17). Once the effect has reached the level where it is apparent or can be identified, it is identified in an animal study as evidence of a pathological effect and in human medicine as a “case,” for which a diagnosis is made.

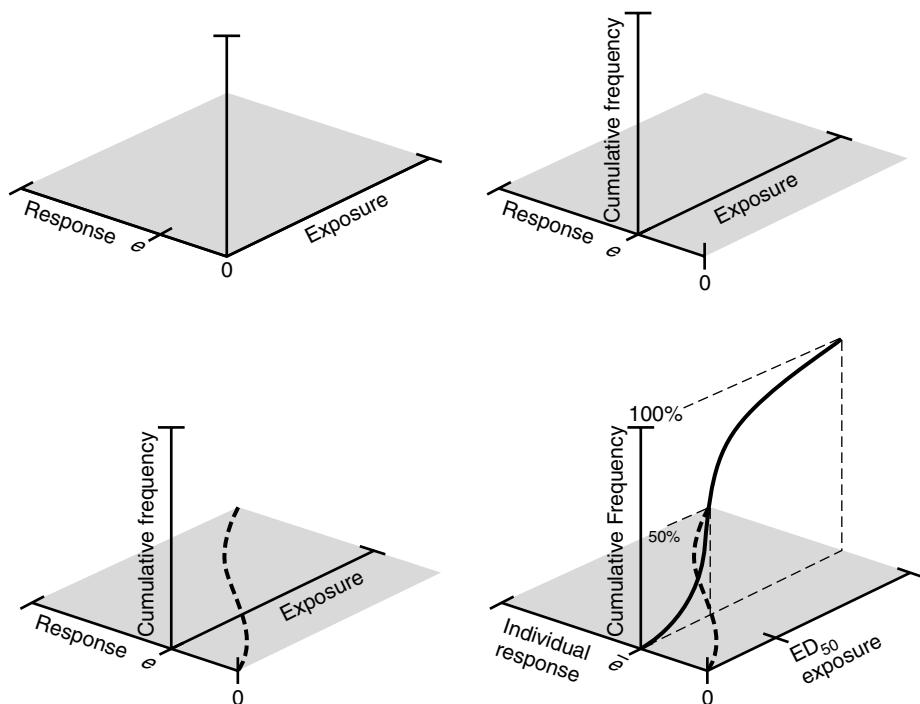
This gives rise to another type of dose- or exposure-response relationship, which might be termed the “clinical” exposure-response relationship (see Figures 2.14 and 2.18). At a given level of exposure—often referred to clinically (but colloquially, not rigorously) as a “threshold” at an effective dose (shown as  $e$  in Figure 2.18)—one can usually expect a given constellation or syndrome of symptoms and signs (a “toxicodrome,” which may or may not be very specific) to accompany significant toxicity. This clinical exposure-response relationship depends importantly on the strength of the host defenses of the individual (which can be very variable) and whether the individual has an acquired condition or genetically determined phenotype that renders him or her more susceptible than others. In a given exposure situation, one person may show one symptom and another a different symptom, based on personal susceptibility. At relatively low levels of lead toxicity, for example, some patients show elevated uric acid levels because of reduced renal clearance, but most do not. The clinical exposure-response relationship, to sum up, describes the first appearance of recognizable effects and then gives a diagnosis for those signs and symptoms, proportionate to exposure.



**Figure 2.18.** Multiple toxicological exposure-response relationships appear clinically to form a “stepladder” of symptoms and syndromes with increasing effective exposure ( $e$ ) for each effect; this is the clinical exposure-response relationship. Adapted from Guidotti TL, Exposure to hazard and individual risk: When occupational medicine gets personal. *Journal of Occupational Medicine*, 1988; 30: 570–577. Reprinted with permission.

Recognition of an adverse effect and the application of a diagnosis to the condition, even if it is provisional, mean that the person who has the diagnosis can be counted as a “case” in epidemiological studies. Epidemiologists formulate “case definitions,” which are working criteria for subjects who are known or thought likely to have the disease or condition of interest. To be recognized as a case implies that the subject has indicators of the disease in the form of symptoms, signs, and positive clinical tests. This logically means that the level of exposure has to have been sufficient to produce this magnitude of effects in the toxicological exposure-response relationship. To be recognized, however, the disorder must be experimentally or clinically detectable. This removes from study many types of response that cannot be directly measured and that are usually considered “subclinical” or “adaptive responses.” Other methods, such as the use of biological indicators or genetic, genomic, or molecular epidemiology are required for that purpose.

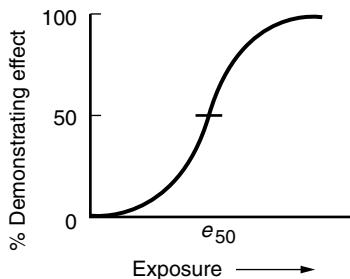
The “epidemiological” exposure-response relationship relates exposure levels to the frequency of the response in a population, not the magnitude of the response or, usually, the severity. (If an epidemiologist is interested in severity, there will be an analysis of indicators that show degree of severity in a disease using subsidiary case definitions formulated for that purpose, but epidemiology does not assess the magnitude of the exposure in a toxicological sense.)



**Figure 2.19.** Plotted on three dimensions, the epidemiological exposure-response relationship is built on the toxicological exposure-response relationship at or exceeding the apparent threshold ( $e$ ) for detection of disease; it is the count of cases that reach this point in the toxicological exposure-response curve.

Given a case definition, the epidemiologist does a “nose count” of cases of the disease among exposed subjects. Plotted in three dimensions (see Figure 2.19), the toxicological exposure-response relationship can be visualized as a plane, and the epidemiological exposure-response relationship can be visualized on a plane perpendicular to it, situated at the magnitude of effect that is consistent with expression of the disease and recognition of a case. The epidemiological exposure-response relationship, to sum up, describes the frequency of recognizable effects, cases, and outcomes proportionate to exposure (see Figures 2.14 and 2.20).

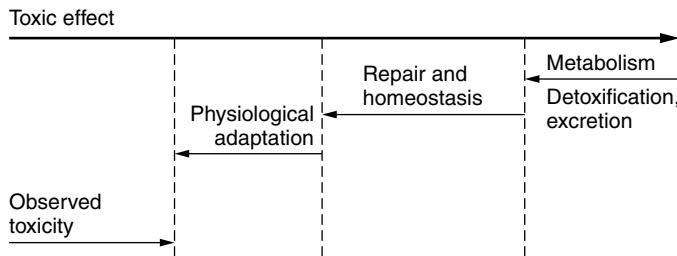
The cumulative count of cases at increasing levels of exposure will also be sigmoid in shape (see Figure 2.20). This is because the under-



**Figure 2.20.** The epidemiological exposure-response relationship. Adapted from Guidotti TL, Exposure to hazard and individual risk: When occupational medicine gets personal. *Journal of Occupational Medicine*, 1988; 30: 570–577. Reprinted with permission.

lying distribution of susceptibility is usually more-or-less normal in the population, and a cumulative plot of a bell-shaped curve is sigmoid. The first few subjects to demonstrate an effect sufficient to be counted as cases will be the most susceptible to the exposure. The few who are least susceptible will be the last to show such an effect and to be counted. The great majority of subjects will fall somewhere in between. “Biological variation” is a general term for variation in biological traits, and susceptibility to many toxic agents—such as the irritant effects of formaldehyde on mucous membranes (throat, eyes)—follows the roughly normal distribution associated with biological variation.

In addition, there are individual, defined susceptibility states present among subjects in the population, which may be genetic or acquired. Genetic states may include conditions such as atopy (and susceptibility to asthma), defects in DNA-repair mechanisms (and therefore cancer risk), antiprotease deficiency, and differences in metabolism of xenobiotics. Acquired characteristics may include conditions such as preexisting disease (such as renal insufficiency), lifestyle factors (such as alcohol intake, which predisposes a subject to chemical injury to the liver), and differences in metabolism of xenobiotics due to smoking or therapeutic drug use. Subjects with these conditions constitute subgroups within the main population. Usually, there is roughly normal distribution of susceptibility within the subgroup itself, but the subgroup as a whole is at greater or lesser



**Figure 2.21.** Toxicological effects are opposed in an intact organism, so the observed effect is the final outcome of a complicated process.

risk than the population as a whole. Most subgroups, except for atopy, are relatively small, and so the overall curve for the population still looks sigmoid.

Toxic effects are not unopposed. There are mechanisms to resist the effects of toxic agents, including induction of metabolism, repair mechanisms, homeostasis, and physiological adaptation (see Figure 2.21). As a result, the observed toxicity in an intact laboratory animal or in the uncompromised human subject is usually somewhat less than might be inferred from *in vitro* studies in the laboratory. Sometimes, in experimental systems such as the nude rat and knockout mice, compensatory and adaptive mechanisms are intentionally removed or impaired so that a toxic effect can be seen more clearly. In animal experiments, interspecies differences in these factors have to be taken into account in extrapolating risk to human subjects. The factors that oppose or offset the toxic effect in human beings are colloquially called “resistance.”

## The Kidney

The kidney is both a major organ of excretion and an important target organ for toxicity. This subsection addresses the kidney as a target organ.

The kidney has a huge functional reserve and can sustain great damage before insufficiency becomes evident. Damage within the kidney tends to occur where blood flow is marginal, where there is

a concentration of molecular target, where xenobiotics are bound, where there are differences in permeability, where xenobiotics are concentrated, and where there is high oxygen demand.

The kidney also has a tendency, manifest in acute renal failure, to follow common pathways of failure once impairment reaches a certain point. For example, when there is an abrupt decrease in the glomerular filtration rate, or increased vascular resistance resulting in low flow, the tubular structure of the kidney starts to fall apart. This results in leakage back into the kidney, cellular debris that obstructs the tubule, cell membrane disintegration, loss of integrins and other proteins that hold cells together, induction of "heat shock" proteins as a nonspecific response to cell stress, and inflammation. Compensation is difficult once the kidney has reached that stage, and the result may be permanent renal failure. This is all familiar to physicians in the pathophysiology of acute renal failure. This is a rare toxicological problem in OEM but is known to be caused by several metals (especially cadmium and mercury), arsine (secondary to massive hemolysis), halogenated alkane hydrocarbons (especially tetrachloroethane, tetrachloromethane [also known as carbon tetrachloride], and 2-chloroethanol [also known as ethylene chlorohydrin]), pentachoropenol, and some nonhalogenated hydrocarbons (including dioxane, not to be confused with dioxin), and may be associated with high exposure levels in occupational settings. Elemental yellow phosphorus is a very potent nephrotoxin that can cause acute kidney and liver failure. Agents and trauma that cause rhabdomyolysis may result in acute renal failure.

Certain structures in the kidney are very susceptible to damage. The epithelium of the glomerulus is easily damaged by cadmium. The basement membrane of the glomerulus is also vulnerable and is the point where immune complexes deposit, causing a proliferative and membranous glomerulonephritis (associated with mercuric chloride and hydrocarbons) and many autoimmune reactions take place. The proximal tubule is an especially vulnerable structure, especially in the part called S<sub>1</sub> (or *pars recta*), because

it is a binding site for many xenobiotics (especially metals, such as lead) and halogenated alkanes (many of them solvents), because xenobiotics are resorbed there, because there are transport mechanisms (such as the one involving metallothionein and cadmium) that bring xenobiotics into the cell against a gradient, because its cells are metabolically active with biotransforming enzymes, because it experiences a high concentration of filtered xenobiotics and their products, and because it has a concentration of metabolizing enzymes in the brush border. The renal papilla is at risk because xenobiotics are concentrated by the countercurrent effect in the loop of Henle, osmotic load is high, and the oxygen tension is low, although this is primarily a problem with medication-related toxicity.

Xenobiotics that are particularly toxic to the kidney in the context of occupational and environmental medicine include metals and solvents, including halogenated hydrocarbons, and possibly some nonhalogenated hydrocarbons at high levels of exposure.

Metals that are more likely to be associated with chronic nephrotoxicity include lead, mercury, uranium, and especially cadmium. Lead induces a nephropathy at the level of the proximal tubule that is often subclinical but that leads to reduced clearance of uric acid and the clinical expression of “saturnine” gout.

Ethylene glycol, most commonly found in antifreeze, is metabolized to oxalate in the kidney, resulting in severe, refractory metabolic acidosis, potentially fatal toxicity, and renal failure unless treated by hemodialysis; the chemical is sweet to the taste and is associated with many inadvertent child and pet exposures from ingestion.

Melamine is a monomer feedstock for melamine formaldehyde resin, which consists of an alternating carbon-nitrogen heterocyclic six-member ring with three amine groups. By itself, melamine has very low toxicity but can cause kidney stones at very high levels of ingestion. Cyanuric acid is a structurally similar molecule with double-bonded oxygen in the same places as the amine groups in melamine, and it is a contaminant after sloppy melamine production but even lower in toxicity as an

individual agent. In combination with cyanuric acid, melamine may form propagating complexes in the distal (not proximal) tubule that result in calculi, irritation, and obstruction. This combined toxicity came to light initially as a result of contaminated pet food and subsequently was responsible for an epidemic of kidney disease, including deaths, in China as a result of adulteration of milk powder used in infant formula. Melamine was added to these products as an illegal way to increase the nitrogen and therefore falsify the apparent protein content on assays for quality assurance. It is an extraordinary example of an interactive effect in nephrotoxicity. There is at present no evidence for occupational toxicity from this cause.

Chloro- and other halogenated alkanes and alkenes are selectively toxic to the proximal tubule, sparing other structures. Trichloromethane (chloroform), in particular, is metabolized by cytochrome oxidase (CYP2B1 and CYP2E1) to phosgene, a very toxic chemical more familiar as an inhaled hazard. This is thought to be a mechanism for carcinogenicity in the kidney as well as nephropathy.

The kidney can also be involved in systemic toxicity, such as beryllium disease, which resembles sarcoidosis, and hemolysis due to arsine toxicity, which releases free hemoglobin and may cause acute kidney failure.

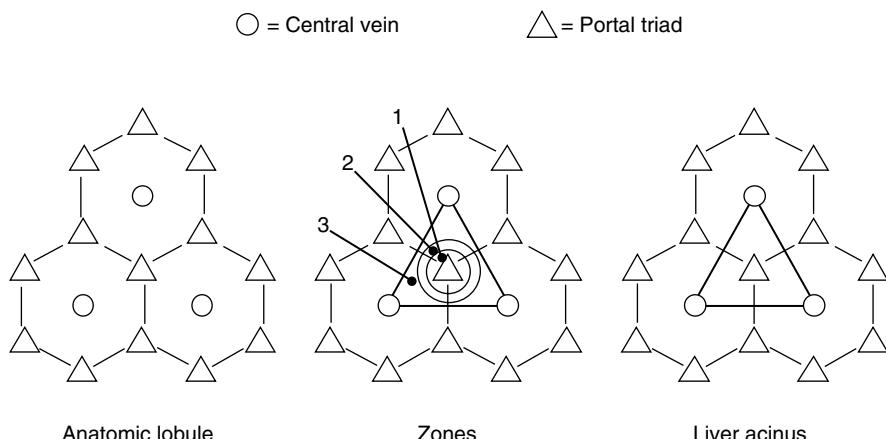
Animal experiments involving nephrotoxicity have to be interpreted with a knowledge of the physiological differences among species. The male rat (but not the female) makes a nonspecific carrier protein called  $\alpha_{2U}$ -globulin that binds with xenobiotics, including lead, and facilitates transport into the cell. Animals with this carrier protein are at risk for nephropathy and cancer after experimental exposure to many common organic compounds, such as gasoline, d-limonene, and lindane, a mostly withdrawn pesticide used for lice and scabies. Human beings do not have this carrier protein and as a result do not share the risk. This is an important example of the difficulties of extrapolating from animal species to human beings.

## The Liver

Like the kidney, the liver has a huge reserve but it also has a unique ability to regenerate itself after injury. This is fortunate because the liver is a particularly vulnerable organ to toxicity from ingested xenobiotics. It is the first organ encountered after portal venous drainage from the gastrointestinal tract. Because of its critical role in metabolic control and synthesis, it is exceptionally rich in biotransforming enzymes, rendering it vulnerable to bioactivated toxic agents. Because of its bulk, its blood supply from the relatively oxygen-poor portal system, and only modest perfusion by the better-oxygenated blood from hepatic arteries, parts of the liver are markedly oxygen deficient compared to other organs. It is rich in binding proteins and lipid and so is a major depot for xenobiotics.

The liver is functionally divided into acini centered on the portal triad (where the portal vein, the hepatic artery, and the bile duct run together in close proximity; see Figure 2.22). (Classical histology and pathology of the liver focused on the lobule, which was centered on the central vein, but this is an anatomic or architectural feature that does not reflect physiology or blood flow.) The acinus is divided into three zones, characterized by distance away from the portal triad and toward the central hepatic vein. The oxygen tension at the tissue level decreases from Zone 1, which is periportal and very active in glycogen and glucose metabolism, to Zone 3, around the hepatic vein, which is rich in cytochrome oxidase and MFO, which are operating in an environment of oxygen deficiency. Xenobiotics that are bioactivated, that place a high metabolic demand on the hepatocyte, that produce free radicals, that require NADH and NADPH (especially alcohol dehydrogenase), and that are stored in the liver have a disproportionate effect on Zone 3. Zone 3 is also the histological location involved in centrilobular necrosis.

The liver can be damaged in many ways. Some xenobiotics act directly on the hepatocyte, such as phosphorus. Some are metabolized to ultimate active agents and cause damage locally. (For example, allyl alcohol,  $\text{H}_2\text{C}=\text{CH}-\text{CH}_2\text{OH}$ , is metabolized to acrolein,



**Figure 2.22.** The hepatic acinus, a functional concept of the architecture of the liver.

$\text{H}_2\text{C}=\text{CH}-\text{CHO}$ , which is highly toxic and causes periportal necrosis, making it specifically toxic to Zone 1.) Other xenobiotics have indirect effects, such as steroid hormones. Immune responses are less common in OEM than as an adverse effect in pharmacology.

Halogenated hydrocarbons at levels relevant to their use as solvents are exceptionally toxic to the liver because they form free radicals during the dehalogenation reaction. Their toxicity is proportionate to the degree of substitution, so that tetrachloromethane, also known as carbon tetrachloride, and tetrachloroethylene are the most toxic of the group, followed by trichloroethylene.

Whatever the cause of the injury, the liver has a limited number of ways to express the damage. Transient mild hepatocellular injury may be evident in transaminase levels, often called a “chemical hepatitis.” Severe hepatocellular injury may occur at high levels of exposure with nitropropane and trinitrotoluene (TNT, better known as an explosive but also an important chemical feedstock) and with nitrosamines. This may be associated with arsenic or halogenated (particularly brominated) hydrocarbons. Fatty liver may follow moderately severe or prolonged hepatitis or may be a direct effect of some agents, such as hydrazine. Severe injury (as might occur with phosphorus or large amounts of selenium) may result in necrosis. Fibrosis

may result, although cirrhosis as a result of chemical exposure is rare. Cholestasis can occur as a direct effect in the case of methylene dianiline, which occurred in one prominent outbreak in the United Kingdom in the 1960s. The liver may also be affected by systemic disease, such as beryllium disease.

Damage to and malignancy of bile ducts, while unusual, is of special concern and has a history in occupational medicine. Angiosarcoma (or “hemangiosarcoma” in the pathology literature) is a very rare malignancy in humans (more common in dogs), arising from vascular tissue (blood or lymph vessels). It is associated with two important xenobiotics of concern in OEM—vinyl chloride and arsenic. The discovery that vinyl chloride was associated, at high levels of exposure, with hepatic hemangiosarcomas was first made in the 1960s by two community-based physicians who saw workers from the same plant in Kentucky.

Evaluation of hepatotoxicity rests on the usual clinical and laboratory methods for evaluating liver function (transaminase levels, bilirubin, serum proteins, coagulation) and for diagnosis (including biopsy and markers such as antimitochondrial antibody). A “chemical hepatitis” resembles an alcoholic hepatitis and can proceed to cirrhosis, although this is uncommon. These tests are neither sensitive nor specific and are good examples of the limitations of clinical testing in evaluating early disease in occupational and environmental toxicology.

## Gastrointestinal Tract

Ingested toxic agents affect the gastrointestinal system in many ways, but few of them are of direct relevance in OEM. These are summarized.

Vomiting can occur by several means. It occurs as a consequence of local gastrointestinal tract stimulation from a wide array of agents including soaps, detergents, solvents, metals (including arsenic and thallium), and toxins associated with several types of food poisoning. It may occur when the central chemoreceptor zone or vomiting center in the fourth ventricle of the brain is stimulated by the chemical. It may occur when there is a central perception of disgust or in

response to a noxious odor or taste. By itself, vomiting is a symptom and an adaptive response, although a distressing one, and not necessarily a toxic effect.

Xenobiotics that are strong acids, strong bases, or chemically reactive, such as phenol, may produce severe tissue destruction or liquefaction necrosis. Alkali with a pH of 11.5 to 12 and acids with a pH of 2 produce significant corrosive injury, which may heal with scarring and esophageal stricture. Less damaging agents may contact the mucosal or other cells of the GI tract, producing erosions, or a superficial ulceration of the mucosa can be widespread or local. Alkali are found in many commercial products, such as household and industrial cleaners, dishwasher soaps and drain openers, and low-phosphate detergents. Factors affecting the degree of tissue injury or destruction include the amount ingested, the duration of contact with tissue, concentration, pH, physical form (granular product is more likely to injure the buccal mucosa, whereas liquid is more likely to injure the esophagus), and buffering capacity. These situations are most often encountered in household incidents involving unsupervised pets or children or suicide attempts but may also occur when chemicals are stored in beverage bottles or are unlabeled or mislabeled.

Lesser ingestion with strong irritants may produce local erosions or ulcers. Erosions are due to focal necrosis of the epithelium and are usually restricted to the superficial layers. Ulcers are deeper lesions extending beyond the mucosa and penetrating into the adjacent tissue layers. Chronic irritation, especially of the esophagus and stomach, can produce proliferative lesions, which can become malignant.

Several pharmacological agents produce GI tract irritation by interfering with the gastric mucosal barrier, most notably the nonsteroidal anti-inflammatory drugs. Metals and iron can cause severe injury to the GI tract, especially the stomach.

The pancreas is uniquely susceptible to high concentrations of ethanol, which can induce an acute inflammatory response characterized by release of amylase into the circulation. If severe, or if this evolves into chronic pancreatitis, it may lead to autodigestion by exocrine enzymes and ultimately to secondary endocrine dysfunction.

Gastrointestinal function can be affected by interaction with cellular receptors. Stimulation of cholinergic muscarinic receptors by agents such as cholinesterase inhibitors (organophosphate pesticides and carbamates) and nicotine can lead to an increase in motility and secretions of the GI tract. This process can lead to symptoms of abdominal pain, cramps, and diarrhea. Similarly, the administration of drugs that block cholinergic muscarinic receptors (for example, atropine, tricyclic antidepressants, opiates, and sedative hypnotic medications) can slow motility and lead to constipation. This is an important part of the symptomatology of cholinergic crisis and the toxicity of the organophosphate insecticides.

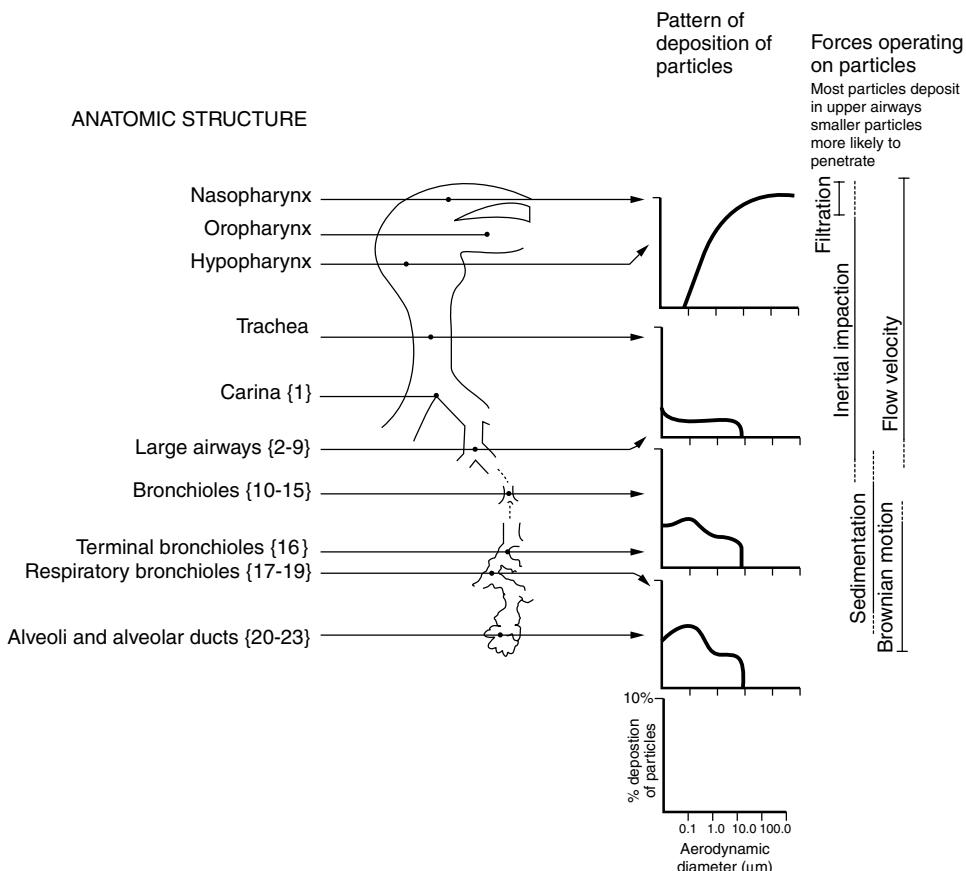
Cancers of the gastrointestinal tract are associated with some occupational and environmental hazards but are mostly associated with lifestyle-related exposures. Oral cavity cancers have been associated with cigarettes, alcohol, and chewing tobacco or snuff or betel-nut quid (popular in parts of Asia). Esophageal cancer has been related to the use of alcohol and nitrosamines, possibly chewing betel nut, and a history of ingestion of alkaline corrosive agents, most often lye. Gastric neoplasms have declined in incidence over the past half-century and have been associated with nitrate ingestion and the formation of nitrosamines. Persons with atrophic gastritis and hypochlorhydria are at greater risk. Colon cancer has been associated with several risk factors, including asbestos, radiation exposure, limited physical activity, and nitrosamines as well as lifestyle factors such as dietary fat intake and low fiber intake. Cancer of the rectum shares some risk factors with colon cancer overall, but also has distinct characteristics possibly related to sexually transmitted infections, chronic inflammation, and cigarette smoking.

## INHALATION TOXICOLOGY

Inhalation toxicology is a distinct branch of toxicology that is particularly critical to occupational and environmental physicians. The lung is a key portal of entry for occupational and environmental hazards. The lung is also a particularly vulnerable target organ of concern

in occupational and environment medicine. Because so many hazards are airborne, inhalation toxicology provides the basis for understanding clinical effects and for evaluating the degree of hazard posed by many toxic agents. An understanding of inhalation toxicology is essential to understanding airborne occupational hazards, ambient air pollution, indoor air quality, the establishing of air quality standards, smoking, and many other issues.

The lung is deceptively simple in basic structure, comprising 300 million alveoli and airways that branch twenty-three times (“generations”). (Figure 2.23 provides a simplified view of lung



**Figure 2.23.** Schematic architecture of the lung, with particle deposition patterns.

architecture, showing patterns of particle deposition.) Gas exchange begins at generation 17 with the respiratory bronchioles. The basic architecture is fixed at birth and subsequent increases in lung size occur due to expansion of alveoli and increasing size of airways. Lung size and function can be reduced by many factors during childhood, including adenovirus infection and passive exposure to cigarette smoke. The velocity of flow of air through the lungs varies considerably and is important in determining particle deposition. The cross-sectional area of the airways is actually narrowest at the level of the “small airways”—the middle-central level of airways, beyond the major airways but well before the peripheral airways, before they branch into bronchioles. This is where resistance to flow is greatest; where the velocity of airflow is highest; where obstruction of fixed airways, as in emphysema, obstructs airflow; and where reactivity of airways, as in asthma, is most likely to obstruct airflow. Beyond the small airways, the cross-sectional area of the airways is much greater, resistance falls, and the velocity of airflow drops so low that air is almost still and gas exchange is by diffusion rather than bulk flow.

The lung is a highly vulnerable organ, open to attack from a variety of airborne agents. An impressive array of defense mechanisms stand guard to protect the lung against toxic gases, particles, aerosolized liquids, viruses, and bacteria. The relative importance of the many defense mechanisms in the respiratory tract probably varies with the individual host and certainly varies with the nature of the threat. These defense mechanisms are remarkably efficient, but they may be overcome even in the normal host by an excessive exposure or by factors that inhibit their effectiveness.

Two general types of exposure are encountered in inhalation toxicology: aerosols (which may be liquid- or solid-phase particles suspended in air) and gases. In binary or complex exposure situations, gases may absorb onto the surface of particles, substantially changing their behavior and toxicologic implications. (An important example of such a binary system is the  $\text{SO}_x$ -particulate complex described in Chapter 12.) These three types of exposure—*aerosols, gases, and gases*

absorbed onto particles—are discussed below, in general terms. The diseases associated with them are discussed in detail in Chapter 17.

Host factors also play an important role in the penetration and deposition of both gases and particles. Tidal volume and breath holding affect the inhalation of toxic agents over short periods of time. Panting with short, shallow breaths tends to reduce penetration of gas and deposition of particles, whereas deep inspirations enhance their penetration and breath holding increases the deposition of inhaled particles. When agents are irritating or have a noticeable odor, the breathing pattern of the subject typically changes to a pant. Reflex mechanisms within the lung may make these adjustments involuntarily and induce bronchoconstriction and coughing. The actual pattern of particle deposition or gas penetration during an acute exposure event may therefore be quite complicated.

## Aerosols

Aerosols are suspensions of particles suspended in air that are stable enough to remain airborne long enough to be inhaled or to move around as an air mass. The particles may be solids or liquids, but must be small enough to remain suspended in the atmosphere for transport if exposure is to occur. Larger and denser particles settle rapidly and never reach the lung. The most commonly encountered aerosol is smoke, in which most of the particles are solid.

The size of small particles, which will be discussed in detail below, is the factor that most influences their behavior in air and in the lung. Other important characteristics include chemical composition, particle, geometry, surface area (a reflection of size and geometry), electrostatic charge, coagulability (tendency to aggregate, in part a function of electrostatic charge), hygroscopy (affinity for water), inertia (partly a function of mass), and solubility (when deposited in aqueous solution, as in the lung or on mucosa, and particularly when taken up by a phagocyte). The deposition of a particle in the lung depends on its size. Once deposited in the lung, such particles may induce an inflammatory reaction, fibrosis, or delayed effects, such as

carcinogenesis. Particles consisting of relatively soluble materials, such as lead, may dissolve slowly (particularly after phagocytosis by macrophages) and release their toxic contents.

Particles can be classified into four general types:

- spheroidal solid particles
- fibers
- droplets
- bioaerosols

Spheroidal particles are characterized principally by their composition and size. They behave more or less like spheres aerodynamically but are rarely truly spherical in shape (except for synthetic albumin and plastic or latex beads used for research and diagnostic purposes). They are usually highly irregular in shape but not grossly elongated.

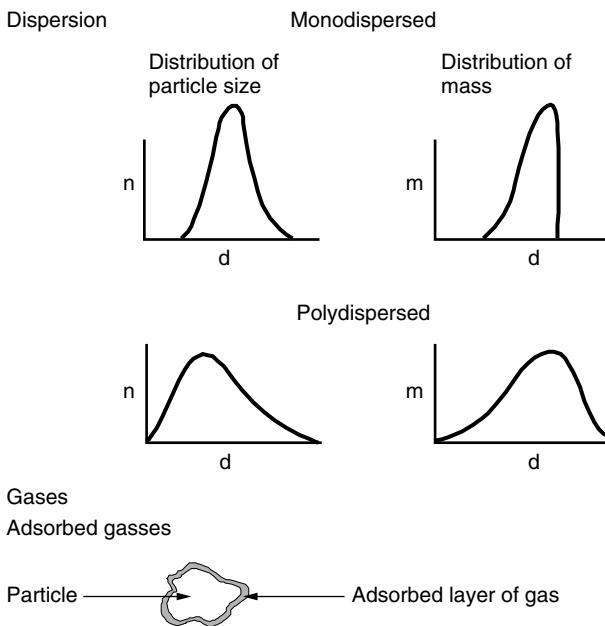
Fibers are elongated in shape and are described by their “aspect ratio” (the ratio of greatest to least geometric diameter, in other words “length” to “width”). Fibers are formally defined as having an aspect ratio of 3:1 or greater (as by the Environmental Protection Agency) but for some purposes, a working definition of 5:1 is used (for example, as a criterion in counting fibers under electron microscopy for purposes of regulation). Asbestos, fibrous glass, and many organic particles assume a fibrous shape and have an aspect ratio typically much greater for length than 5:1. Talc, on the other hand, is “stubby” in shape and has an aspect ratio closer to 3:1. Mineral fibers are often difficult to characterize fully because they may split, fracture, or aggregate, forming fibrils and complicated, irregular structures. That is why the assessment of fibers—for example in the Asbestos Hazard Emergency Response Act (AHERA) protocol for counting asbestos fibers in the air of schools— involves counting rules and restrictive operational definitions of what is considered to be a fiber.

Droplets are particles of liquid phase, approaching a truly spherical shape because of surface tension. With time, evaporation from the

surface of a droplet of water or a volatile liquid causes it to shrink to a small core of liquid with concentrated solute, called a nucleus. These nuclei may contain a high concentration of dissolved or suspended material in the droplet and, because they are so small, remain suspended in air for long periods. Droplets become more concentrated with time and droplet nuclei are important in the transmission of tuberculosis, influenza, and probably other respiratory disorders.

Bioaerosols are aerosols consisting of particles of biological origin and tend to have a size range between 0.5 and 100  $\mu\text{m}$ . They are the result of biological activity, such as mold spores, pollen, dust-mite feces, animal dander, and human skin flakes. They may be infectious (infectious particles, droplets or droplet nuclei, zoonotic agents such as dust carrying Q fever), noninfectious but toxic (such as environmental endotoxins, fungal glucans), allergenic (including molds, dust-mite feces, cockroach parts, aerosolized cat urine or saliva), or, exceptionally, agents of aggression (such as weapons of bioterrorism or biological warfare).

Size varies within natural and anthropogenic aerosols. An aerosol that has uniform particle size is said to be “monodispersed” and never occurs except by design. A monodispersed aerosol, in which all particles are of uniform size or within a very narrow range, is technically difficult to make and precisely monodispersed systems are thought to be achievable only in the absence of gravity. Natural aerosols such as windblown soil, dust, or smoke; bioaerosols; droplet clouds produced by sneezing; fog and sea spray; and aerosols arising from normal human activity, such as combustion products, sawdust, mechanical abrasion, particulate air pollution, and house dust show variation in size and so are called “polydispersed.” In a polydispersed aerosol, the majority of particles are small in size but carry little mass. The greatest mass is in the larger and heavier, but far fewer, particles in the distribution. The size and mass distributions will be mirror images of each other, and each shows a typical “log-normal” distribution, familiar in most exposure-assessment situations. Formally, the variance in particle size is described by the geometric standard deviation ( $\sigma_g$ ) because the distribution is geometric, not normal. Figure 2.24 illus-



**Figure 2.24.** Characteristics of aerosols that are important in inhalation toxicology.

trates and summarizes the characteristic features of aerosols pertinent to inhalation toxicology.

Size, which is determined in practice by sedimentation characteristics, is the major characteristic that determines the behavior of a particle in the respiratory tract. The “aerodynamic diameter” of a particle is a calculated measurement (formally the “mass median aerodynamic diameter”) that reflects what the diameter would be if the particle were a perfect sphere falling at the same velocity, corrected for density. Size in this case is measured not by geometric dimensions but by sedimentation velocity. The mass median aerodynamic diameter for the aerosol is used to describe polydispersed aerosols. Another way of describing particles in an aerosol is by “cut size,” a method used mostly for dusts in occupational hygiene and in describing particulate air pollution. The cut size specifies the maximum aerodynamic diameter of a dust—the size that defines the upper end of the modal distribution. For example, PM<sub>10</sub>, in air pollution, refers to particulate

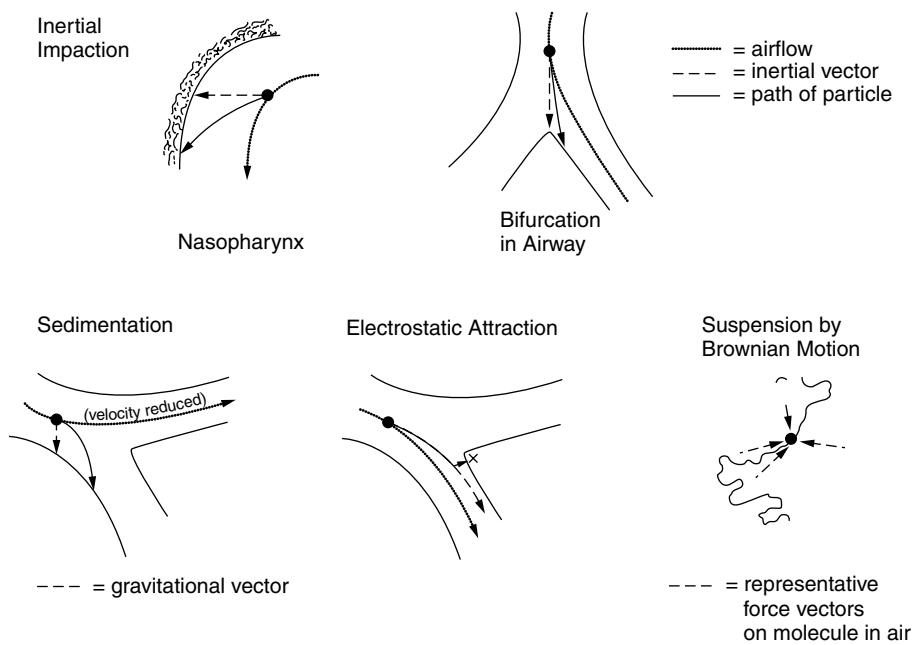
matter 10  $\mu\text{m}$  or smaller in size.  $\text{PM}_{2.5}$  refers to particulate matter 2.5  $\mu\text{m}$  or smaller. The cut size is used when the distribution of size in an aerosol is known in general terms and one is trying to characterize a peak in the distribution.

The anatomy of the airways, which are a series of bifurcations, also affects the deposition of dust particles (see Figure 2.23). Secondary flows and eddies result in local, nonuniform flow and slowing of forward velocity, and permit foci of increased deposition at the bifurcation. Thus, the behavior of particles in the 0.5 to 2.0  $\mu\text{m}$  range is governed by boundary layer and laminar flow phenomena, resulting in disproportionate deposition at bronchial and bronchiolar bifurcations. These phenomena dictate fairly uniform deposition for particles below 1.0  $\mu\text{m}$  but increasing nonuniformity of deposition with increasing respiratory velocity for larger particles, up to 5  $\mu\text{m}$ .

The respiratory tract presents a complicated picture, as illustrated in Figure 2.23. Beyond the carina, the lower respiratory tract bifurcates into twenty-three generations of airways before ending at the acinus, a unit consisting of alveoli, alveolar ducts, and respiratory bronchioles served by a single terminal bronchiole. If a particle is inhaled through the nose, it is filtered through the hairs of the nose and must negotiate the eddies and tortuosities of the nasal turbinates. As a particle passes posteriorly, it must negotiate the curve at the hypopharynx, where if it is too large or dense it will impinge due to inertia. It must then negotiate each of the twenty-three progressively smaller bifurcation points before reaching the alveoli. Beyond the large airways, flow velocity abruptly drops off, leaving the particle suspended in relatively slow-moving air. If it is too large or dense at each point, the particle will sediment by gravity or impinge through inertia at these points and become trapped on the mucosa of the airway. As a result, many particles of even the smallest diameter are deposited before they reach the alveoli. Indeed, the internal angle of airways bifurcation is the characteristic site of carcinoma *in situ* of the lung precisely because that is where particles are most likely to deposit. Figure 2.24 illustrates these important forces.

At each level of the lower respiratory tract, the particles that deposit most efficiently become smaller and smaller in size. Only particles in the respirable range penetrate to the alveoli, where they become entrapped, leading to disorders such as the pneumoconioses. However, most particles, even in the respirable range, will deposit before reaching the alveoli. Figure 2.25 presents a summary of the pattern of particle deposition at various levels of the lower respiratory tract. (The curve for the terminal bronchiole is hypothetical.) This figure is intended to show that particles deposit with different efficiency at different positions along the lower respiratory tract. This difference can be manipulated for clinical applications. For example, by producing droplet aerosols greater than  $10 \mu\text{m}$  in diameter, deposition of bronchodilator-containing droplets in the larger airways can be enhanced. This is the principle behind nebulized medication.

Figure 2.25 illustrates the processes by which particles are sorted by size as they progress through the airway to the deep lung. Particles above



**Figure 2.25.** Particle behavior in the airway.

100  $\mu\text{m}$  aerodynamic diameter tend to be deposited in the nose and larger airways due to inertial impaction. They simply cannot negotiate the bends and the curves through the nose, pharynx, and at the carina. Particles this large tend to stay in the upper respiratory tract. They may be irritating and may cause local effects in the nose, nasopharynx, trachea, and sinuses, but they will not penetrate more deeply.

For smaller particles, particle deposition can also be viewed as a race between the tendency of the particle to settle by gravity and its tendency to remain aloft as it is carried along in the airway. The velocity of sedimentation of a particle is expressed as

$$V_s = \rho dg / 18\eta$$

where  $d$  is the aerodynamic diameter of the particle,  $g$  is the vector of gravitational acceleration,  $\eta$  is the viscosity of air, and  $\rho$  is the density of the particle relative to water. Thus, the velocity of sedimentation depends not only on particle size but also on density. The distribution of particles from a mixed aerosol can be envisioned as a sedimentation gradient, with smaller and lighter particles sedimenting more slowly and also penetrating more deeply into the lung.

Particles above 10  $\mu\text{m}$  become deposited on the wall of the airways and at bronchial bifurcations due to inertia or gravitational sedimentation. (Inertial impaction is the reason that bronchogenic carcinoma usually starts at the inside bifurcation of airways.) Fibers tend to behave somewhat differently than spheroidal particles in air because a leading or trailing end of the fiber may intercept an airway, depositing it sooner than it otherwise would be deposited.

Particles with an aerodynamic diameter less than 10  $\mu\text{m}$  but greater than 5  $\mu\text{m}$  attain sufficient inertia to negotiate the changes in direction of the larger airway bifurcations but eventually impact and become deposited on the wall of the smaller airways due to inertial impaction or gravitational sedimentation.

Particulate matter of 1 to 5  $\mu\text{m}$  deposits more efficiently at the alveolar level. The particles in the midrange are most likely to

penetrate to the alveoli and, once there, to sediment and to deposit on the airspace wall. This so-called "respiratory range" between about 5  $\mu\text{m}$  and 0.5  $\mu\text{m}$  is the range of most efficient deposition in the alveoli and is therefore of great importance in the pathogenesis of lung disease, such as the pneumoconioses.

Particles down to 0.5  $\mu\text{m}$  are small enough to negotiate changes in direction of airflow but may settle out by sedimentation once they reach the low-velocity area of the peripheral airways and alveoli. They are most likely to remain in the alveoli.

Fine particulate ("little particles") matter of less than 0.1  $\mu\text{m}$  in diameter may either remain airborne—kept aloft by the familiar phenomenon of Brownian motion—and be exhaled without depositing in the lung. Some particulate may be pushed into mucosa by diffusion, and others may be attracted to the mucosal wall by the Coulomb effect of electrostatic attraction, because particles so small usually carry a positive charge and the mucosal wall carries a slight negative charge. However, the majority will remain suspended, and the aerosol behaves in many ways like a gas.

## Gases

An airborne gas is a gas dissolved or suspended in the gases of the atmosphere and carried along by air movement. The term includes vapors, which usually result from recent evaporation from the surface of a liquid.

Airborne gases are somewhat simpler than particles as physical systems, but they are by no means simple. The characteristics of gases that determine their toxicological behavior include chemical reactivity, density, and solubility. Of these, solubility is the characteristic of greatest practical importance in determining penetration into the lower respiratory tract. Gases, once inhaled, have no mechanical barrier to penetration into the lower respiratory tract, but are subject to rapid removal by crossing the alveolar barrier, entering the blood, and being carried away into the systemic circulation. Solubility therefore plays a key role in determining

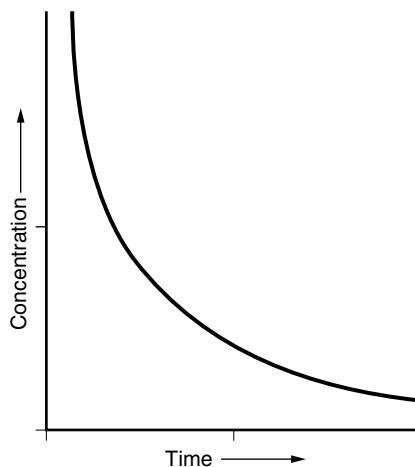
where injury may take place. Chemical reactivity is most important in determining the mechanisms of toxicity and producing the actual toxic effect. Density is of practical importance only in unusual situations, such as in hyperbaric environments and deep ocean diving, in physiological research, or in the helium test for small airways disease.

Solubility, in this context, is defined by the ease of dissolution of a gas in water. As gases that are soluble in water—such as ammonia, sulfur dioxide, or chlorine—come into contact with the moist surface of the mucosa, they dissolve and are removed from the airstream. Gases that are relatively insoluble in water, such as ozone or nitrogen dioxide, penetrate more deeply and reach the alveoli with less reduction in concentration along the way. More-soluble gases may never reach the alveoli in significant concentrations except in cases of overwhelming exposure, deep and sustained inspiration, or when carried adsorbed onto an aerosol.

The nasopharynx and the trachea are highly efficient at removing soluble gases and absorbing large fractions of airborne agents such as ozone, nitrogen dioxide, sulfur dioxide, and ammonia before they reach the lower respiratory tract. The upper airway can therefore be thought of as a scrubber removing a large fraction of soluble gas before it reaches the lower respiratory tract. Soluble gases may have local effects, such as bronchoconstriction, cough, and bronchitis, but usually do not cause deeper lung injury except at very high concentrations, which might occur if exposure takes place in a confined space. Significant alveolar effects, such as pulmonary edema (discussed in Chapter 17), require deep penetration of relatively high concentrations of most toxic gases. Insoluble gases are therefore much more likely to cause acute alveolar injury than soluble gases. Soluble gases do occasionally cause acute alveolar injury when concentrations are high and duration is relatively long, a situation most likely to occur in a confined-space incident.

Toxic gases that penetrate to the alveoli have free access to the pulmonary capillary bed and may diffuse directly into the bloodstream. Some gases do so without direct damage to the lung. Carbon monoxide, for example, has no effect on the lung itself.

Some gases have a cumulative effect over short periods, such that the same endpoint may occur with short periods of high exposure or long periods of lower exposure. “Haber’s Law”— $C \times t \approx K$ —is a commonly used expression in toxicology, indicating that the product of concentration ( $C$ ) and duration of exposure ( $t$ ) for a toxic gas yields a constant effect, usually taken to be lethality or pulmonary edema. It assumes a relatively predictable, roughly proportionate effect for concentration and time. Figure 2.26 illustrates Haber’s Law. This formula is the basis for the “time-weighted average” used in occupational exposure standards (such as OSHA Permissible Exposure Limits, PELs) and guidelines (such as the American Conference of Governmental Industrial Hygienists’ Threshold Limit Values, TLVs) (see Chapter 7). In reality, Haber’s Law applies strictly only to a few gases, such as phosgene. It does not accurately reflect the toxicological behavior of most gases, and the true behavior of many gases, such as hydrogen sulfide, deviates so far from Haber’s Law that its application is misleading. (Toxicity of hydrogen sulfide is driven overwhelmingly by concentration and only secondarily by duration of exposure.) The “law” was named after Fritz Haber, the great German chemist and toxicologist, who is said to have



**Figure 2.26.** Haber’s so-called law for toxic gases:  $C \times t \approx K$ , where  $C$  = concentration,  $t$  = time, and  $K$  = constant for a given toxic effect, usually death or pulmonary edema.

been outraged that such an unreliable equation was named after him. As a practical matter, the unreliability of Haber's Law at levels of exposure encountered in occupational settings is probably not a major drawback, but it does suggest that the toxicological logic behind TWAs and TLVs is questionable.

### Aerosol-Gas Mixtures

Situations in which potentially toxic gases are generated together with aerosols are common in occupational medicine. Examples include open fires, automotive and diesel exhaust, and cigarette smoke. Many gases and some liquids adsorb onto the surface of particles, forming a thin film coating the particle, as illustrated in Figure 2.16. As a result, soluble gases may be carried much more deeply into the lower respiratory tract than their solubility would otherwise allow them to penetrate. This is a particularly serious issue in the case of combined exposure to sulfur dioxide and particulate air pollution, as discussed in Chapter 12. The "sulfur dioxide–particulate complex" was a major cause of symptoms during air pollution episodes in the middle of the twentieth century, when industrial air pollution had a reducing chemistry and particulate levels were very high.

Likewise, droplets in the liquid phase can carry dissolved gases.

### Host Defenses

The lung is a highly vulnerable organ, open to attack from a variety of airborne agents: toxic gases, particles, aerosolized liquids, viruses, and bacteria. An impressive array of defense mechanisms stand guard to protect the lung, but they may be overcome even in the normal host. Table 2.3 lists some of the more important host defense mechanisms of the lung.

The nose and nasopharynx may be likened to the sentinel and gatekeeper. There are at least eight identifiable ways in which nasal mechanisms protect the lung. The nasopharynx is an effective heat-transfer device that appears to warm inspired air before

it enters the lungs and to conserve heat from expired air. It is equipped with an extensive vascular erectile mechanism under autonomic control for this purpose. Humidification of inspired air may protect the lower respiratory airways from desiccation. Filtration through nasal hairs and inertial impaction (while negotiating the 90° curve into the hypopharynx) removes the largest particles and thus reduces the burden on the mucociliary net. Mucociliary transport may clear the particle by delivering it to the oropharynx, where it may be expectorated or swallowed, or by delivering it to the accumulations of lymphoid tissue at Waldeyer's ring (at the back of the throat, including the tonsils), where it may be attacked immunologically. The oropharynx is rich in secretory IgA, which protects the airway from viruses and bacteria. The nasopharynx and the trachea, as noted above, are highly efficient "scrubbing" mechanisms for soluble toxic gases, absorbing large fractions of airborne agents before they reach the lower respiratory tract—up to 70 percent of even less-soluble gases such as ozone and nitrogen dioxide, and 90 percent or more of sulfur dioxide and ammonia.

Of great importance—and often overlooked—is the role of the nose in initiating avoidance behavior. The detection of noxious odors is clearly an evolutionary adaptive response; although many contemporary airborne threats, such as carbon monoxide, are odorless. The perception of an odor may also initiate physiological responses, such as nausea and emesis, as well as behavioral responses and anxiety. Odors may also trigger conditioned behavior, such as when an odor is associated with a bad experience or fear of injury due to toxicity.

The unique pulmonary defense against foreign bodies and accumulated mucus is appropriately pneumatic: coughing. This is a complex reflex that generates intrathoracic pressures of great magnitude to forcefully expel the offending irritant. The same mechanism of inertial impaction that operates in the nasopharynx also works in the bronchial tree to filter out particles larger than about 3  $\mu\text{m}$  in aerodynamic diameter.

**Table 2.3.** Principal Host Defenses of the Respiratory Tract

Nose and upper airway

- Thermal regulation
- Humidification
- Filtration
- Inertial impaction
- Mucociliary transport
- Delivery to lymphatic tissue
- Absorption (soluble gases)
- Olfaction (avoidance behavior)
- Sneeze
- Cough (in response to upper airway obstruction)

Lower respiratory tract

- Cough
- Inertial impaction
- Mucociliary transport
- Detoxification/neutralization
- Immunoglobulins
  - IgA—bronchial fluid
  - IgG—alveolar fluid
- Cellular immune functions
  - T-cell mediated cytotoxicity
  - Cytokine-mediated reactions
- Alveolar fluid constituents
  - $\alpha_1$ -Antitrypsin and other protease inhibitors
  - Complement

Alveolar macrophage functions

- Phagocytosis
- Lysosomal digestion
- Myeloperoxidase-peroxide-halide activity
- Secretory activity
  - Myeloperoxidase system
  - Lysosomal enzymes
  - Interferon
  - Complement factors
  - Pyrogen

(Continued)

**Table 2.3.** (*Continued*)

- 
- Cell-to-cell mediators (cytokine)
  - Foreign-body reaction
  - Detoxification
- 

Mucociliary transport is a complex process that plays a key role in the defense of the lungs. Cilia in the human respiratory tract measure about 5–8 μm in length and 0.15–0.3 μm in diameter, and number some 200 per ciliated cell. The characteristic 9 + 2 filament internal structure of the cilium is associated at the base with the cell basal body, and the filaments are firmly inserted into the cellular cytoplasm by “striated rootlets.” The ratio of ciliated cells to goblet cells is about 5:1 in the respiratory epithelium. Ciliary propulsion of mucus is described as “antileptic metachronal” because successive waves of ciliary beats propel fluid in the opposite direction of the wave. The ciliary beats are energy dependent and normally continuous, consisting of a propulsive stroke with the cilium straight followed by a recovery stroke with the cilium bent. The strokes are separated by a resting period and have a time ratio of 1:3, the entire process requiring only about 0.05 second. Surrounding the cilia there is probably a less viscid periciliary fluid, with the characteristic that transmission of energy to it is very efficient; its relaxation time relative to the ciliary beat would be such that the cilia beat on a virtual elastic solid. The overlying mucus, heavy with glycoprotein, is a product of the goblet cell. The periciliary fluid may be a derivative of the alveolar fluid. This transport mechanism is exceedingly efficient and produces mucus velocities estimated to be 100–400 μm/sec.

Immune and immune-related mechanisms are very important in the lung, both locally and, if the offending agents win the initial local skirmish, systemically. Secretory IgA is present in bronchial secretions and IgG in alveolar fluid. Cellular immune functions may be direct, such as cell-mediated cytotoxicity and the secretion of cytotoxic lymphokines, or mediated through other effector cells, particularly alveolar macrophages summoned to battle by chemotactic and

migration-inhibitory factors. (The roles of macrophages will be further discussed below.) Alveolar fluid also contains complement, although less than in serum. Of particular importance, protease inhibitors, chiefly  $\alpha_1$ -antitrypsin, act to prevent unnecessary damage from local tissue-injury proteases released from phagocytes.

Of all the pulmonary defense mechanisms, however, the most versatile is the alveolar macrophage. Phagocytosis and intravacuolar digestion by lysosomal enzymes are probably the best-known functions of the alveolar macrophage. The cell also possesses a myeloperoxidase halide system similar to that of polymorphonuclear leukocytes, which is an effective mechanism for killing bacteria. The macrophage also secretes a variety of substances, including lysosomal enzymes for local extracellular digestion, interferon to protect against viral invasion, components of complement (at least C4 and C2), endogenous pyrogen, and informational molecules (such as interleukins) that coordinate the counterattack with lymphocytes. Alveolar macrophages act against a foreign body or persistent antigen, as in tuberculosis.

These host defenses work remarkably well in the normal host, but they may be overcome, making the lung susceptible to injury and infection.

### *Smoking*

Cigarette smoking, aside from its potential carcinogenic and cardiovascular properties, is perhaps the most obviously preventable process that impairs the defenses of the respiratory tract. Smoke contains a number of acutely toxic substances, among them nitrogen dioxide, acrolein, carbon monoxide, cyanide, organic acids, and aldehydes (see Chapter 11). Cigarette smoke is ciliostatic, although filtration to remove the gaseous from the particulate phase greatly reduces toxicity to cilia. Chronic bronchitis, induced by smoking, has been associated with a much higher content of smaller-sized glycoproteins in mucus than in normal controls. The implications of this change for mucociliary clearance are unclear, but decreasing the relaxation time of the mucus layer might inhibit ciliary propulsion. The mucus layer

(or “gel phase” of the mucus layer) is like a rubber conveyor belt—anything that either greatly increases or greatly reduces its elasticity will reduce its effectiveness. The effect of cigarette smoke on the alveolar macrophage is dramatic: the cells are more numerous; they lose many of their surface characteristics; and they are drastically impaired in phagocytosis, responsiveness to migration inhibitory factor, and protein synthesis. Smoking also depresses serum antibody responses to infection and may inhibit lymphocyte proliferation in lymphoid tissues. It induces aryl hydrocarbon hydroxylase activity in macrophages, potentially interfering with detoxification mechanisms. In brief, cigarette smoking has a wide variety of detrimental effects on the defenses of the respiratory tract.

### ***Air Pollution***

Air pollution is discussed in detail in Chapter 12. The role of air pollution in compromising host defenses is much less clear than the role of smoking. The constituents of air pollution vary among communities, but where industrial and automotive emissions are mixed, particulates usually include hydrocarbon, sulfur dioxide, gaseous hydrocarbons (including irritating aldehydes), ozone, nitrogen dioxide, and carbon monoxide. Other important pollutants, collectively called “air toxics,” may be present in communities with certain industries.

Sulfur dioxide, ozone, and nitrogen dioxide share a number of the same characteristics as toxic gases. They are ciliotoxic, especially when adsorbed onto particulates, and they result in ciliary degeneration as well as stasis. Macrophage functions—including phagocytosis, interferon production, and intracellular killing—are reduced by somewhat higher concentrations of the gases, particularly nitrogen dioxide. A growing body of experimental work has documented bacterial clearance, reduced bactericidal activity, and increased clinical severity of both bacterial and viral lower-respiratory tract infections after exposure to relatively low concentrations of nitrogen dioxide and sulfur dioxide. The public health implications of such depressed host resistance may be

significant. There is particularly strong evidence to suggest an increased morbidity from respiratory diseases in children living where the ambient levels of nitrogen dioxide are high. In short, air pollution effects are exceedingly complicated and have subtle influences on host defenses. They are particularly important, however, because unlike cigarette smoking, they result from an involuntary exposure of an entire, unselected population, sick and well, and their control is a matter of public policy rather than individual choice in habits.

### *Alcohol*

The pharmacological effects of alcohol may significantly compromise host defenses in the respiratory tract, compatible with the clinical observation that alcoholics are at increased risk of pneumonia. Aside from the risk of aspiration during stupor or seizure, alcohol inhibits ciliary motility and tends to dehydrate mucus. Acute alcohol excess also has a mildly immunosuppressive effect on alcoholic individuals, reducing delayed hypersensitivity and antibody response to an antigen; polymorphonuclear leukocyte chemotaxis is also impaired. Thus, chronic alcohol abuse may well impair the host defenses of the respiratory tract in many ways.

### *Other Chemical Agents*

Cilia are remarkably resistant to changes in pH, maintaining motility at pH values as low as 5.0, although motion is impaired below 5.7. On the other hand, cilia tolerate alkalinity as high as pH 10.0 with no detectable impairment. Ciliary motility is unsynchronized but not stopped by local anesthetics such as lidocaine and procaine, whereas tetracaine or dibucaine, and inhaled anesthetics such as halothane, do suppress ciliary motility. Cocaine not only inhibits ciliary motility but may also inhibit the secretion of mucus. Opiates, barbiturates, and possibly anticholinergics, given systemically, appear to decrease mucociliary transport.

### *Cold Temperature*

Except for frank hypothermia, temperature extremes do not appear to have much effect on the respiratory tract defenses. Over a wide range of temperatures of inspired air, there is little significant change in nasal mucus flow or airflow resistance. Although cold air impinging on tracheal cilia is known to reduce ciliary wave frequency, the nasopharynx has an elaborate mechanism for air warming that protects deeper structures. Despite elaborate studies, no association has been observed between the attack rate, illness duration, viral shedding, or clinical severity of experimentally induced rhinovirus infections and temperature of the environment. The virus itself is heat sensitive, however, and its proliferation may be inhibited by warming the hypopharynx directly or, adaptively, by fever. Cold temperatures also provoke reactivity in airways.

### *Viral Infections*

Viral infections may have a devastating, but fortunately reversible, effect on respiratory defenses. Rhinoviruses, adenoviruses, enteroviruses, parainfluenza, and respiratory syncytial viruses appear to have similar effects. Mucus production is inhibited, and particulate clearance by mucociliary transport is impaired. Extensive necrosis and desquamation of epithelial ciliated and goblet cells follow. Viral infection may cause defects in intracellular killing capacity in alveolar macrophages, a plausible mechanism for bacterial superinfection. Even upper respiratory tract infections may have significant effects on the lower respiratory tract.

## **Experimental Inhalation Toxicology**

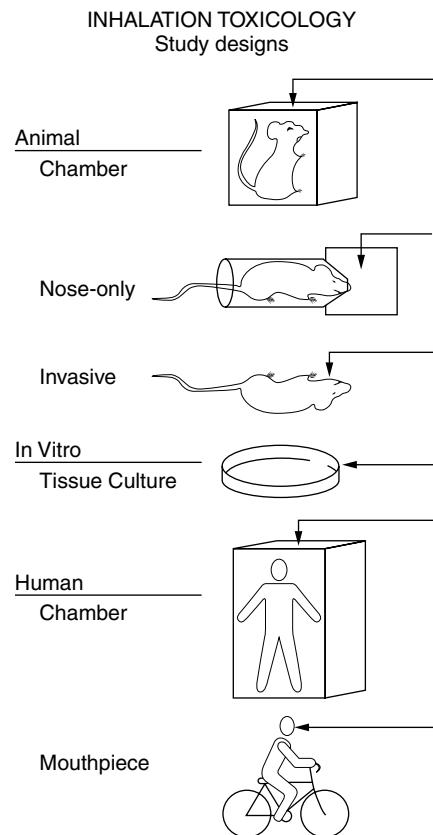
Inhalation toxicology tends to be more complicated and difficult than other areas of experimental toxicology. The cost of experiments is often higher, and fewer laboratories work in the field because of high infrastructure costs, particularly for chamber experiments.

Particles represent a substantial technological challenge in inhalation toxicology, and the means of generating consistent particles and maintaining their concentrations accurately has been a major engineering achievement.

Inhalation toxicology protocols for animal experiments may involve single, acute exposures, subchronic exposures (thirteen weeks or less), or chronic exposures lasting one year to a lifetime. The test material may include particles, gas, antigens, and radionuclides. Exposure may be static, placing the animal in an atmosphere for a short period of time; recirculating, returning the same atmosphere to the exposure chamber for a time; or dynamic. For safety, human exposure always requires dynamic exposure systems. Dynamic exposure is the most technically difficult but most scientifically valid if it is important to have a consistent exposure at a known level for a relatively prolonged period. Dynamic exposure systems involve generating the atmosphere, which may involve a storage vessel (such as a gas cylinder); continual synthesis of the gas (for example, using electrical equipment to produce a constant supply of ozone); generation of a particle; treatment of the atmosphere (as by mixing, or allowing elements of the atmosphere to react together, as in experiments on air pollution); dilution; delivery to an exposure device; recovery and exhaust; and monitoring to ensure constant concentration and quality assurance.

Inhalation toxicology uses a limited number of experimental methods (illustrated in Figure 2.27):

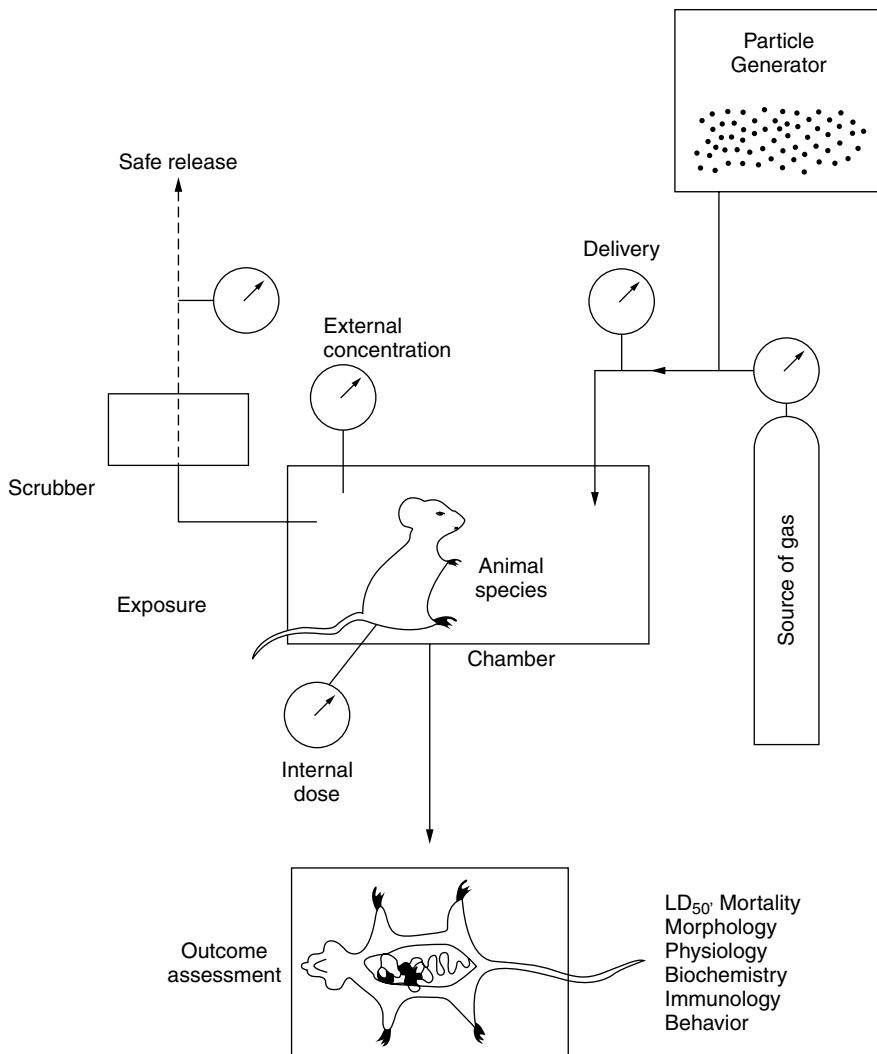
- chamber studies, involving animals
- nose-only exposure, involving animals
- invasive airway access, by intubation
- intratracheal instillation
- *in vitro* studies, usually involving tissue culture
- chamber studies involving human exposure
- mouthpiece studies involving human exposure



**Figure 2.27.** Experimental systems in inhalation toxicology.

Chamber studies involving animals are technically difficult. A simplistic chamber model is illustrated in Figure 2.28, which only hints at the technical complexity involved. They require a large chamber constructed of a nonreactive material (usually stainless steel) fitted with shelves or cage-holders to accommodate wire cages through which air can circulate freely. The chamber must be meticulously clean. The atmosphere going into the chamber must be carefully monitored to ensure that the concentration of gases and particles is constant and matches the protocol. The concentration of the atmosphere inside the chamber must also be monitored to ensure

that animals are breathing what they are supposed to be breathing. If the gases or particles being studied are toxic or if emissions might exceed air pollution emissions standards, the exhaust needs to pass through a scrubber or other cleaning device. Inside the chamber, the ventilation needs to be carefully balanced and mixed in order to



**Figure 2.28.** Simplified schema of a chamber system for inhalation studies.

prevent pockets of poor air circulation, currents in which air flows in and out again without mixing, or laminar flow against walls. The rate of air exchange has to be brisk in order to prevent accumulation of carbon monoxide and to keep the atmosphere constant, without creating a wind inside the chamber. Animals in the chamber tend to modify the exposure by urinating, which releases ammonia and may neutralize acidic exposures, or by burying their noses in their fur. Keeping animal exposure chambers operating requires specially trained personnel and is very expensive. Almost all inhalation toxicology studies today in chambers use rats or mice. Chambers are used for long-term exposure studies and for gases that are hazardous or irritating.

Nose-only exposure uses the “nose cone,” a Plexiglas cylinder with a truncated cone that is open at the end and that can be sealed at the other end. The rat is placed into the cylinder and pushed forward so that its nose sticks out through the opening. The hind end is sealed, often in such a way that the tail can be drawn through a hole so that blood can be drawn. The cylinder is placed into a port on a bench-top exposure chamber that can accommodate several units. The atmosphere is delivered into the exposure chamber, and the rats, which are obligate nose breathers, inhale a uniform atmosphere from the chamber. Nose-only experiments have many advantages. They are well tolerated by rats, very convenient, and much less expensive than chamber studies, and the tail vein is accessible to draw blood; and because the cylinder can be made effectively air tight, pulmonary function studies can be done on the rat by plethysmography. Nose-only systems are especially convenient for studying particulate air pollution and mixtures. Nose-only exposure cannot be used for long exposure periods, but it can be used for repeated short-term exposures.

Invasive access through intubation is unusual in animal studies today but is sometimes used when the experimental species is larger than a rodent.

Intratracheal instillation involves delivering an agent of interest, usually a suspension of particles into the trachea through a cannula by syringe. It is a useful means of studying the lung's response to a

number of agents on a preliminary basis before proceeding to inhalation experiments or dosing the lung for carcinogenesis studies. Intra-tracheal instillation is relatively easy and inexpensive. It has a potential drawback in that the distribution of particles in the lung that results is not the same as would occur by inhalation.

As lung biology becomes more sophisticated, *in vitro* experiments have become more popular and informative. These include tissue culture of lung cells and exposure to gases in the atmosphere above the culture.

Human chamber studies are less common than in the past, when they played a major role in elucidating the health effects of air pollution. These chambers are large enough to accommodate human subjects comfortably for hours. In the past, subjects lived in them for days, but such studies have become prohibitively expensive, too difficult to conduct with contemporary standards of human subjects protection, and unattractive to volunteers. Because of the need to protect human subjects, exposures are limited to atmospheres equal or only slightly exceeding what would be encountered in daily life or within occupational or environmental exposure standards. Human chamber studies usually involve serial measurements of physiological parameters, such as pulmonary function and exercise tolerance. The heyday of such studies was in the late 1970s and early 1980s, when they played an essential role in discovering the health effects of air pollution.

Mouthpiece studies are human studies in which the subject breathes through a mouthpiece. Usually this is done while on a treadmill or riding a bicycle ergometer, so that exercise tolerance and other physiological measurements can be made. The major limitation of such studies is that they can only be performed for a limited time.

## **CARCINOGENESIS**

Carcinogenesis is a multistep stochastic process, meaning that it depends on a set of contingent random probabilities at each step along the way. At each step in the sequence there is a finite probability

of events leading to the next step. Chemical carcinogenesis is thus a stochastic, or probabilistic, process, functioning like a roulette wheel or radioactive decay, and not a certain prediction based on chemical structure and properties. The process is not stepwise and certain, and there are no “Koch’s postulates” to be satisfied, as in infectious disease. In any one individual, an exposure may increase the odds of getting cancer, but the fact of exposure does not make it certain in absolute terms that this will happen.

OEM physicians regularly deal with issues regarding the risk of chemical exposures in causing cancer. A basic understanding of the process of carcinogenesis helps to clarify issues related to occupational cancer or suspected carcinogens (see Chapters 10 and 17). This section is not a primer on carcinogenesis or molecular oncology. It is intended to highlight important concepts that apply primarily to OEM.

OEM is sometimes criticized for excessive concern with cancer as an outcome. Actually, that concern is rational. Cancer is, after injury, the second leading cause of both potential years of life lost and disability-adjusted years of life lost in the United States and Canada, because it affects a disproportionate number of younger people in years of their lives when they would not be otherwise disabled, compared to heart disease. The public also has an understandable dread of cancer (see Chapter 7), and regulatory policy reflects this.

Concern over the risk of cancer drives much of the regulatory process for setting standards for allowable occupational exposure to chemicals and various forms of radiation. The perception of the public is that there has been an epidemic of cancer linked conclusively with environmental exposure to chemicals. Nevertheless, except for lung cancer resulting from exposure to cigarette smoke and occupational exposure to asbestos, there is no substantial overall epidemic of cancer in Western societies once age is taken into account. People are living much longer, in general, and surviving to get cancer later in life. Certain types of cancer are certainly on the increase, such as pancreatic cancer and melanoma, but

it is equally true that other types of cancer are becoming much less frequent, such as stomach.

The most common lethal cancer in North America is lung cancer, which is now declining among men but until recently was still increasing among women as a result of changing smoking habits. Despite these generally encouraging trends, mortality from a small number of cancers appears to be increasing among people fifty-five years of age and older: brain, melanoma, breast, kidney, non-Hodgkin lymphoma, and myeloma. These counter-trends are explained by most epidemiologists as reflecting improved diagnosis, a greater willingness on the part of physicians to attribute cause of death in the elderly, and more complete reporting, but a minority of epidemiologists think that rising cancer rates at advanced ages represent the cumulative effect of lifetime exposure to chemical carcinogens in modern life. Even so, age-specific cancer rates for the general, younger population do not show these anomalies, and they do not disprove the general observation that cancer death rates are going down, not up, when adjusted for age. The overall death rate (mortality) from cancer is also slowly but steadily improving for most cancers. Lung cancer has been exceptionally refractory in that regard.

One major reason for the public perception of a general cancer epidemic is that the population is aging and people are surviving to older ages, especially women. Most cancers have a higher incidence with increasing age, with the risk usually increasing exponentially with age. Cancer has therefore become much more visible to people as the population comes to include more individuals in the older age groups. Many more people are themselves afflicted or have close friends or family members who have or have had cancer. The average North American now has a one in four (males) or five (females) chance of dying of some form of cancer. An estimated 5 million Americans alive today either have or have had a diagnosis of cancer, mostly of breast, colon, uterus, prostate, bladder, and lung.

Another reason for the public's perception of a cancer epidemic is that cancer is given much attention from the media, which regularly covers stories of suspected cancer risks, "unexplained" cancer

clusters, and medical breakthroughs relating to cancer. Cancer is also a topic of conversation today, but it was shunned in years past. Scientific knowledge of cancer and the basic biology of carcinogenesis have increased rapidly and have been widely publicized, raising expectations for breakthroughs in practice. The public's expectations became greatly inflated by the so-called "War on Cancer" of the early 1970s and subsequent media attention. The public is understandably impatient for quick results in treatment. There have been many premature announcements regarding the efficacy of innovative treatment that implied a cure for cancer was just around the corner and had only to be found through a concerted effort such as that of the space program. This turned out to be overly optimistic, and the public is to some extent disappointed, despite real progress in cancer therapy and in understanding the mechanisms of carcinogenesis.

### **Chemical Carcinogens**

The authoritative world body in recognizing carcinogens is the International Agency for Research on Cancer (IARC). IARC periodically reviews the scientific evidence for carcinogenicity in animals and humans of individual chemicals, mixtures, certain industrial processes, and sometimes occupations as a whole (in 2007, firefighters). IARC issues reports ("monographs") summarizing the evidence, which is often very complex and incomplete, and categorizing the agents according to a simple scheme based on the weight of evidence:

- Group 1: sufficient evidence for carcinogenic activity in humans
- Group 2A: limited evidence for carcinogenic activity in humans, probable human carcinogen based on animal evidence
- Group 2B: limited evidence for carcinogenic activity in humans, possible human carcinogen based on animal studies
- Group 3: insufficient evidence for carcinogenic activity
- Group 4: evidence for an absence of carcinogenic activity

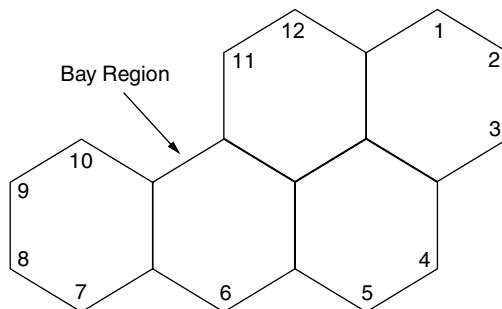
Because the IARC system is fundamentally conservative, the findings of other agencies are often used to supplement IARC's list with probable or possible carcinogens that merit regulation. These include the Environmental Protection Agency and the California Department of Health, which has been especially proactive in identifying putative carcinogens.

Many chemicals are identified as possible carcinogens because their chemical structure resembles that of known carcinogens and may act in the same way. At present, it is clear that certain classes of organic chemicals include many members that are carcinogens:

- Polycyclic aromatic hydrocarbons (specifically, benzo(a)pyrene, chrysene, benzo(a)anthracene, benzo(b)fluoranthene, dibenzo(a,h)anthracene, dibenzopyrene, and 5-methylchrysene). These have in common a concave, electron-rich area in their structure called the "bay region" (Figure 2.29), which reacts more readily with electron-poor regions of DNA bases (see Chapter 10).
- *N*-nitrosamines
- Mycotoxins (especially aflatoxin)
- Aromatic amines
- Alkylating agents (which mimic radiation and may be alkylating or arylalkylating)
- Metals (arsenic, cadmium, chromium VI, nickel subsulfide, and probably beryllium and lead)

Some of these carcinogens are specific as to site:

- Bladder: benzidine, 2-naphthylamine, 4-aminobiphenyl, PAHs, the rubber industry (the specific carcinogen has never been identified but is thought to be one or more nitrosamines)
- Blood (benzene, specific for acute myelogenous leukemia)
- Larynx (sulfuric acid mist)
- Lung: arsenic, asbestos, beryllium, bis chloromethyl ether and chloromethyl ether, cadmium, chromium VI, nickel, PAHs,



**Figure 2.29.** Nucleophilic “bay region” of benzo(a)pyrene.

ionizing radiation, sulfuric acid mist, painters, foundry workers, hematite mining

- Nasal cavity: nickel, leather dust, wood dust, isopropanol manufacturing
- Pleura (mesothelioma): asbestos
- Bone:  $\alpha$ -emitting radionuclides, including radium and thorium
- Skin: arsenic, PAHs, mineral oils, ultraviolet radiation, ionizing radiation, and coal hydrogenation

Many of these chemicals are associated with multiple sites, such as arsenic. (This presents a particular challenge to the application of the Hill criteria for assessment of causation in epidemiological studies, as discussed in Chapter 3.) Many occur in the same environments (silica, asbestos, and radon, for example, used to occur regularly together in mines). Other carcinogens are discussed in the context of occupational cancer in Chapter 12.

## Chemical Carcinogenesis

Carcinogenesis is not a straightforward, deterministic process. At each step in the sequence, there is a finite, generally relatively low probability of events leading to the next step. Chemical carcinogenesis is a probabilistic process, like a game of roulette or radioactive decay,

rather than a certain prediction based on chemical structure and properties. In any individual, exposure to a carcinogen may increase the odds of getting cancer, but exposure does not make cancer certain. Chemical carcinogens increase the frequency of cancers in exposed subjects compared to unexposed. They sometimes produce tumors that differ in tissue type from those usually observed among unexposed subjects, but they often only elevate the rate at which more common cancers appear. Similar mechanisms operate for chemical and radiological carcinogens, and the science of chemical carcinogenesis has been greatly informed and influenced by studies on ionizing radiation.

The history of research in carcinogenesis begins with an occupational risk. Ramazzini observed that breast cancer was more common in nuns, the first recognition of an occupational cancer. The recognition that chemicals may cause cancers derives directly from Percivall Pott's initial observation in chimney sweeps in 1775, noted above, which applied to skin. F. H. Hartung and W. Hesse recognized lung cancer in miners in the Schneeberg region of central Europe, in 1879, becoming the first to recognize occupational exposure as a cause of internal cancer. Rehn, in 1895, first identified an association between a specific chemical and cancer by demonstrating that analine dye workers were at risk for bladder cancer. Yamagiwa and Ichikawa, extending the observation of Pott, used coal tar on the skin of rabbits to create the first reproducible carcinogen bioassay in 1915. Sir Ernest Kennaway was the first to isolate an active carcinogen from a mixture in 1930, separating benzo(a)pyrene from coal tar. Wilhelm Hueper studied aniline dye-induced bladder cancer in dogs, producing in 1938 the first animal model of a chemically induced internal cancer. In 1950 Richard Doll demonstrated the association between lung cancer and cigarette smoking, the first carcinogen identified through pure epidemiology. Umberto Saffiotti and colleagues developed the first carcinogenesis bioassay protocols at the National Cancer Institute in 1968, which launched the NCI Carcinogenesis Program, which later, in 1978, became the National Toxicology Program. In 1975 Bruce Ames developed the in vitro mutagenicity assay for genotoxicity presumptively associated with carcinogenicity.

The covalent binding of benzo(a)pyrene to DNA was demonstrated by many investigators around 1976. The revolution in biomedical research since then has vastly deepened and broadened the field, and genomics is quickly revolutionizing it again. As understanding of the basic mechanisms of cancer has improved, concepts of chemical carcinogenesis have also grown more refined. Today, carcinogenesis research focuses on signal transduction and genomics and is rather abstractly removed from OEM practice. Even so, the mechanisms of carcinogenesis help to explain what the OEM physician may observe, inform conclusions on causation, and provide the logic behind risk management strategies.

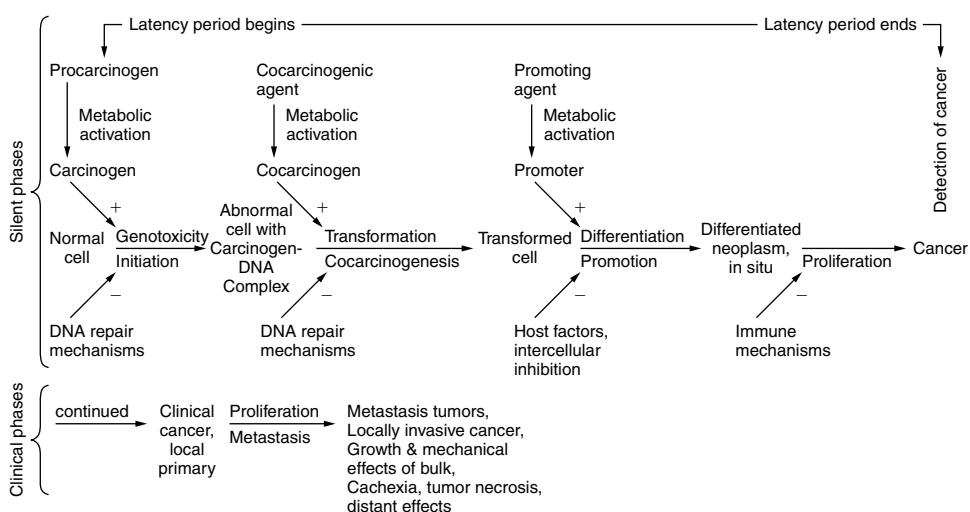
A deeper understanding of the biology of cancer has helped to explain many of the phenomena critical to regulation and control, such as latency periods and cancer promotion. With a better comprehension of the process, it should become possible for society to set standards of exposure that provide greater assurance of protection, and to anticipate problems with newly synthesized or introduced chemicals. It may even be possible to reduce the risk of cancer once exposure to a carcinogen has occurred. When this is done by the administration of a chemical, whether a nutritional constituent or a drug, it is called “chemoprevention.” Considerable interest has arisen in recent years in applying chemoprevention to populations of workers occupationally exposed to carcinogens. Clinical trials were undertaken in the United States in the 1980s to determine the efficacy of cis-retinoic acid (a vitamin A derivative) in blocking steps in the sequence of events leading to certain cancers. Unfortunately, the intervention resulted in a paradoxical increase in cancer risk, forcing the approach to be abandoned. An effective and safe chemical intervention would be invaluable, however.

The induction of cancer by chemical carcinogens involves several steps, which occur in sequence (see Figure 2.30):

1. Delivery of an ultimate carcinogen to the target cell
2. Initiation, which usually involves covalent binding to DNA, producing a somatic mutation

3. Fixation and promotion, in which the mutation is stabilized by mitosis
4. Gene expression, in which the cell is transformed
5. Neoplastic growth and proliferation of a clone
6. Progression, with the development of local effects
7. Metastasis

Each step has a probability of occurring and may be influenced by other exposures. Chemicals may be initiators, acting exclusively or primarily at the state of initiation; co-carcinogens, acting to facilitate the action of an ultimate carcinogen at the time of initiation; or promoters, enhancing the fixation of somatic mutations and increasing the probability of cancer after initiation. Carcinogens may be “incomplete,” meaning that they require a promoter to increase the risk of cancer, or “complete,” meaning that they both initiate and promote carcinogenesis. Initiation and promotion are dose dependent. Initiation may or may not have a threshold (the default assumption in risk management is that it does not), but promotion probably does.



**Figure 2.30.** Simplified sequence of events in carcinogenesis.

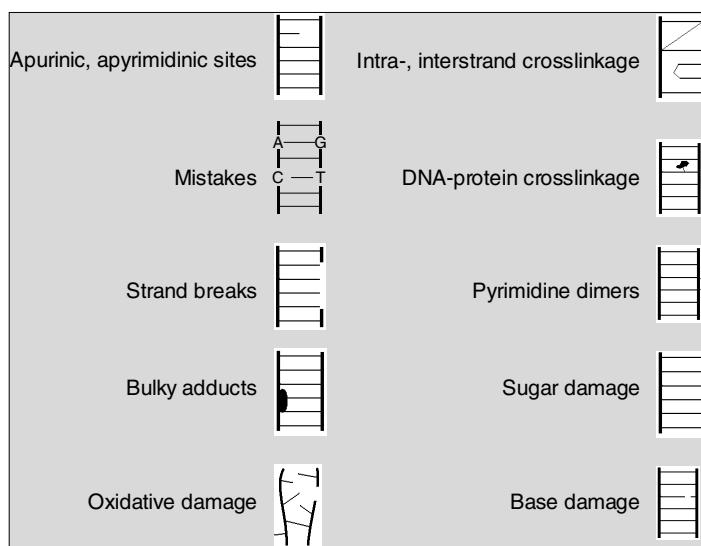
It takes time for a cancer to go through all these steps. The total period from time of first exposure to the detection of the cancer is called the “latency period.” However, this is a clinical or epidemiological definition, not a toxicological one. In fact, one cannot know if the first event in the sequence leading to cancer took place at the time of first exposure or later. The point of detection of a cancer is not fixed either, and will obviously depend on the size of the tumor, where it is located, and whether it causes local signs or complications. From the point of view of toxicology, latency consists of at least four components:

1. Time from first exposure to point of actual initiation
2. Time from initiation through dormancy to promotion
3. Time from promotion to proliferation
4. Time from onset of proliferation to detection

It is known from animal experiments confirmed by epidemiological data that the latency period can be modified. Higher exposure shortens the latency and is one reason why animal cancer bioassays use very high doses in order to induce an excess of cancers within the lifetime of the animal. Latency also appears to be proportional to the age of first exposure, so that exposure in younger years seems to be associated with shorter latency periods. Latency is often estimated at twenty years as a rule of thumb, but this applies only to the common solid tumors of epithelial origin and is highly variably even within this group. Latencies much shorter than this are seen in leukemias, bladder cancer, and some mesotheliomas. Latencies much longer than this are routinely observed in many lymphomas, lung cancer, and most mesotheliomas.

The first step in the process, as outlined above, is the delivery of the ultimate carcinogen to the target tissue. This may involve the biotransformation of a “procarcinogen” to an “ultimate carcinogen” product, as described above in the metabolism of xenobiotics. The ultimate carcinogen typically binds with many macromolecules, mostly to no effect. However, adducts to proteins or hemoglobin may be useful biomarkers for exposure.

Initiation requires interaction of the ultimate carcinogen with DNA. After covalent binding to a DNA base, the adduct this forms may be repaired by DNA endonuclease, restoring the normal genome and releasing DNA adducts (which may also be used as biomarkers). One example of a DNA adduct is the covalent linkage of aminopyrene (AP, a polycyclic aromatic hydrocarbon compound) to the amino group in guanine to form C8-AP-dG, a big, bulky, multi-ringed molecule hanging on to the amino group at the C8 position of a deoxyguanine in the DNA chain. The AP group interferes with the normal function of DNA, both in “unzipping” the double helix and normal transcription. In this case, the bulk of the adduct is the mechanism of DNA damage. The specific type of damage to DNA can take many forms: loss of purines or pyrimidines at a site, transcription errors, strand breaks, inter- and intra-strand cross-links, DNA-protein cross-links, pyrimidine dimmers within a strand of DNA, damage to the sugar backbone of DNA, and damage to bases. Figure 2.31 illustrates the various forms of DNA damage that may occur.

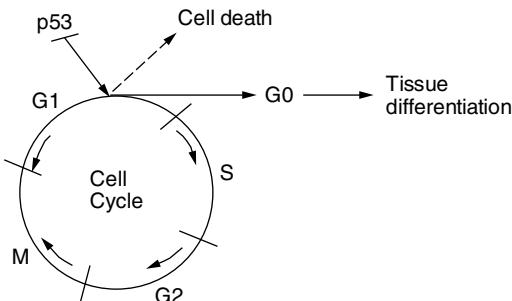


**Figure 2.31.** Patterns of DNA damage associated with genotoxicity and carcinogenesis.

Damage can also occur as the result of oxidation damage or the effect of reactive oxygen species of inflammation on DNA. There are many endogenous sources of damage to DNA as well, so there is also an ongoing background risk of carcinogenesis. Most DNA that is damaged will be repaired by the usual excision-repair mechanism, although double-strand breaks and end-joining are particularly tricky to repair. If the damage to DNA cannot be repaired and interferes with critical functions, the cell will die, or apoptosis (programmed, carefully controlled cell death) will be triggered. A small number of nonlethal DNA adducts or damaged DNA segments remain and cause a stable transcription error. In a small number of cases, DNA will be repaired incorrectly, and if the resulting code is viable, DNA polymerase will replicate the error. The result is a somatic mutation.

A somatic mutation that has the effect of de-repressing a protooncogene may activate oncogene gene expression for processes that result in undifferentiated cell growth. There are a variety of ways that this can happen, depending on the oncogene. Oncogenes are activated, unregulated versions of protooncogenes, which are normal genes encoding for protein kinase and other growth signals. Their gene products stimulate cell growth and are themselves convenient biomarkers of effect. Most appear to be viral in origin but are fully incorporated into the human genome. Single copies or alleles of oncogenes are sufficient to result in malignant transformation of the cell.

At the same time, a second mutation may occur that results in inhibition of tumor suppressor genes, which exist to halt the cell cycle to prevent proliferation and to eliminate cells that have aberrant genetic features. The most common tumor suppressor gene is p53, which stops the cell in mid-cycle (at the G1 phase, see Figure 2.32) and triggers apoptosis to get rid of aberrant cells. This prevents a clone from getting started. The second mutation has the effect of blocking the pathway to apoptosis and prevents the elimination of the aberrant cell. In a sense, it removes the brake on the runaway cell. Thus, the somatic mutation escapes suppression and survives. For this to happen, both alleles of the tumor suppressor gene must be knocked out—not just one, as in the case of oncogene activation.



**Figure 2.32.** Tumor suppressor mechanisms that arrest the cell cycle stop replication of cells with DNA damage, unless mutations occur.

The next phase is promotion. The cell affected by a carcinogen must replicate for the process of carcinogenesis to proceed. Cell division fixes the mutation in daughter cells so that it can be inherited. Certain chemicals called “promoters” induce rapid tissue growth by irritation, necrosis, or hyperplasia and facilitate the fixation of the mutation in daughter cells. The promoter also changes gene expression in ways that facilitate survival of the somatic mutation. Promoters act at the same time or after initiation. They are not usually initiators themselves at the dosage in which they promote carcinogenesis, the major exception being polycyclic aromatic hydrocarbons. Promoters may be exogenous xenobiotics (such as phorbol esters, phenobarbital, dioxins, saccharine), endogenous chemicals (cholic acid), hormones (estrogen, prolactin, thyroxin), or foreign bodies (such as subcutaneous metal or plastic implants). Promoters induce small foci of preneoplastic proliferation around transformed cells, which then have a competitive advantage due to their de-repressed growth.

In vitro, transformed cells change their appearance markedly. The changes are not so obvious *in vivo* but are sufficient for detection by cytopathology if they are sufficiently numerous and aberrant. The transformed cell may lie dormant for many years. There must be very many transformed cells, because in the end, few result in viable cancers. The abnormal cell may remain dormant for a very long time,

contributing the greater part of the latency period before appearance of a clinically evident tumor. Dormancy may end for a number of reasons, such as hormonal stimulation, nutritional changes, stimulation by lymphokines, or transient immunodeficiency, allowing the clone to escape surveillance or suppression.

Next in sequence is the growth of a clone of transformed cells from a single cell, altered in its growth characteristics, through clonal expansion and proliferation to a small focus of *carcinoma in situ*. There are some xenobiotics that may stimulate transformed cells to progress, although this is not proven: arsenic salts, asbestos, benzene, and hydroxyurea have been proposed.

The proliferating cancer cells assume the essential features of a malignancy: unresponsiveness to regulation, loss of contact inhibition, potential for sloughing and migration of cells, and the potential for inducing growth of new nutrient blood vessels (neovascularization). Local invasion and mechanical effects that lead to clinical detection are largely a function of tumor size and the natural propensity to shed cells and therefore are not obvious until the mass of the primary cancer has passed through two or three dozen “doubling times,” which is the length of time it takes to double the number of cancer cells and therefore, approximately, the tumor mass.

Sometimes, medicolegal issues hinge on whether a tumor is new or was present before a certain time. Sometimes, medical experts are tempted to argue that a tumor is a certain age by measuring it on the chest film with a ruler and extrapolating back from the doubling time. Doubling cannot be assessed accurately from an X-ray alone, even a solitary lung cancer on a chest film, because the opacity on a plane surface, the film, is only the projection of a three-dimensional object onto two dimensions. If a tumor mass is perfectly spherical, its mass would be proportional to the radius cubed (to the third power), but the area of the shadow on the film is proportional to the radius squared. Masses are usually irregular, so the growth may not fully project onto the plane of the film. The doubling time is more useful conceptually in explaining why a mass was present for years before it became visible. Doubling time is really an abstraction because there may be cell loss in

the tumor occurring at the same time that cells are proliferating, and local factors, such as insufficient blood supply, may impede tumor growth in some parts of the malignancy and not others.

Tumor doubling times vary from years to months but are rarely only weeks and never days. A tumor that doubles in 100 or 60 days is a very aggressive malignancy. If it takes thirty doubling times and a mass of a billion cells to reach a detectable tumor (which would still be rather small, roughly 1 cm<sup>3</sup> in diameter), a tumor with such a short doubling time would still have been growing, in place, for about eight or five years, respectively. The doubling time of an epithelium-derived carcinoma is measured in months.

Finally, if it escapes immune surveillance and does not compromise its own vascular supply, the tumor mass may have grown sufficiently large to be detectable. It is only at this late phase that screening programs have a role. Less aggressive and more accessible malignancies, such as breast and cervical cancer, are more readily detected by screening. Cancers that are usually aggressive and that metastasize early (such as lung cancer and melanoma) or that are difficult to detect because of their location (such as pancreas and ovary) do not lend themselves to effective management by early detection and treatment. Aggressive screening for cancers that cannot be effectively treated has the net effect of lengthening the time after which the tumor is recognized without affecting the outcome, leading to a false increase in apparent age-specific incidence without a change in cancer-specific mortality. This was the situation with respect to early screening for lung cancer by chest film, before the introduction of spiral CT scanning, which is being evaluated for its potential as a screening tool in high-risk populations, such as smokers and workers occupationally exposed to asbestos. Sometimes, changes in technology lead to detection of increased numbers of tumors overall, the majority of which may not be the cause of cancer-specific mortality. This appears to be the case for intraductal breast cancer *in situ* and for nonaggressive prostate cancer, leading in both cases to initial confusion over whether a true change took place in incidence and mortality; this confusion was

resolved by the recognition that the new screening tools were simply picking up cancers that would otherwise have remained silent. The result is that improved diagnostic modalities may easily result in spurious increases in estimated incidence, without really affecting the frequency of cancers or mortality.

To metastasize, malignant cells must digest or displace the matrix binding them—especially basement membranes—migrate through the degraded tissue, gain access to blood or lymphatic vessels for transport, and be deposited in a tissue favorable to growth. This usually does not occur until the tumor has reached a size of at least 1 cm<sup>3</sup>, but this obviously varies with tissue type. By the time metastasis has taken place, the prognosis for effective treatment for most cancers has deteriorated, and screening programs are not very useful.

### ***Predisposing Factors***

Dietary factors appear to influence cancer risk and may modify risk for cancers of occupational and environmental origin. These include caloric intake; protein deficiency and high-fat, low-carotene and retinoid intake (although supplemental carotene does not appear to be protective); tocopherol intake; selenium deficiency (which reduces glutathione peroxidase levels); and zinc deficiency. Flavonoids appear to be protective.

Cancer risk is a function of interaction between genetic factors and exposure to carcinogenic agents. There are many examples of genetic predisposition to cancer, including the familiar autosomal dominant models of familial polyposis and retinoblastoma, although these examples are not strictly relevant to chemical carcinogenesis. It is presumed that genomic factors play a significant, if not a determinative, role in occupational and environmental cancers following exposure to a carcinogen.

There are many susceptibility states that modify individual risk of cancer. These include the diverse family of DNA repair deficiencies, such as ataxia-telangiectasia, which confers great susceptibility to radiation-induced cancer. Errors in the metabolism of xenobiotics

and immunodeficient states may confer greater risk of cancer. The genetic contribution to cancer risk may be multifactoral: there are at least sixteen genes associated with the risk of lung cancer. It is well known that individuals with one cancer, even a skin cancer, are at elevated risk for a second. Other risk profiles remain to be identified.

It has long been known that certain families are “cancer-prone,” with a much higher risk of epithelial cancers, such as lung cancer, than in the general population. This trait also seems to follow the inherited pattern of susceptibility for chronic obstructive lung disease and other respiratory disorders.

However, specific genetic defects account for only a small fraction of occupational cancers and, by extension, environmentally derived cancers. The general principle remains that genetic predisposition is an important risk factor for some individuals but that for populations as a whole, levels of exposure to the carcinogen are much more important.

### *Epigenetic Carcinogenesis*

Epigenetics is a major topic in genomics today, referring to modifications of DNA, for example by methylation of bases, that accumulate over a lifetime and that may even be heritable, leading to an environmentally induced modification of the genetic code that can be passed on to the next generation. This meaning of “epigenetic” promises great insights for the future in gene expression and toxicology. However, the older use of the term “epigenetic” is what is meant in this subsection.

In this context “epigenetic” refers to the actions of cancer-inducing agents and exposures that do not interact directly with DNA. At least some probably act by producing intracellular free radicals that damage DNA in a nonspecific manner. Others are more obscure in their mechanisms. None are adequately explained by the conventional two-stage model of carcinogenesis, but subsequent refinements in theory will almost certainly result in a unitary model demonstrating a final common mechanism for most cancers. In molecular oncology in general, epigenetic carcinogenesis is an important anom-

aly but not a major practical concern. In occupational cancer studies, it is a major issue because so many important carcinogens act by epigenetic mechanisms.

Epigenetic carcinogens are of particular concern in risk assessment because they cannot be identified by the usual in vitro assays and are therefore likely to be missed in tier-one testing. There are many carcinogens in this category that are important in industry, such as benzene, dioxane (not to be confused with the dioxins), and possibly asbestos. There are several more associated with consumer products (nitriloacetic acid, known also as NTA), medical devices (foreign body implantation), and pharmaceutical agents (hormones). Hormonal induction of cancer is also of great importance in occupational and environmental medicine because of the weakly estrogenic effects of many substituted hydrocarbon compounds, including some pesticides, and as a concern in the environment with respect to “endocrine disruptor” chemicals.

The epigenetic carcinogens of greatest significance to occupational and environmental medicine include the following:

- Asbestos and silica. Direct interaction with DNA does not appear to be the mechanism of action of either carcinogen. Neither asbestos nor silica has been shown to act by genetic means, although both are proven and relatively potent carcinogens. Available evidence suggests that the persistent inflammatory reaction in response to the presence of asbestos fibers may result in the production of intracellular free radicals that damage DNA in target cells and cell-to-cell mediation that affects target cells.
- Foreign body carcinogenesis. Implantation of relatively inert materials, such as plastic prostheses or embedded fibers, is associated with the subsequent development of sarcomas in animals and in humans, particularly in implanted medical devices. The mechanism is thought to be production of reactive free radicals and cellular mediators as a result of the persistent inflammatory response, leading to secondary changes

in nearby target cells. This mechanism may also operate in the induction of mesothelioma, which, although it is associated with inhaled asbestos fibers, can also be induced by implanting a variety of fibers, including cotton, into the pleural space.

- Hormones. Hormone-induced carcinogenesis remains controversial. Estrogens are potentially carcinogenic when used in nonphysiologic applications, particularly the synthetic estrogen diethylstilbestrol, which caused vaginal adenocarcinomas in female children years after it was used by their mothers for prevention of fetal loss. Occupationally, exposure to these chemicals has caused problems in the pharmaceutical industry, including an outbreak of gynecomastia in Puerto Rico. Two uncommon chemicals that resemble thyroid hormones in structure, the herbicide 3-aminotriazole and the synthetic rubber constituent ethylene thiourea, have been associated with thyroid cancers in animals, but an effect in humans has not been demonstrated for either.
- Nitriloacetic acid (NTA). A chelating agent used in a variety of chemical processes in industry and a constituent replacing the more environmentally problematical phosphates in laundry detergents, NTA is a kidney and bladder carcinogen in rodents given in high doses, but it does not react directly with DNA.
- Dioxane. Not to be confused with the dioxins, this common solvent has a low toxicity and has been generally considered acceptably safe for use in the workplace. In unusual and very high exposures in animal experiments, however, it appears to be capable of inducing liver and possibly nasopharyngeal cancers without directly interacting with DNA. Dioxane resembles benzene in structure.
- Benzene. An important industrial chemical in the past as a solvent and today as a constituent of petroleum products and byproduct of hydrocarbon combustion, benzene is conclusively linked with the induction of leukemia and aplastic anemia in humans and various solid neoplasms and lymphomas in animal

tests. Control of benzene exposure has been a major issue in regulatory policy and litigation. The exact mechanism of action for benzene is not known, except that it does not act directly on DNA. This compound is the single most important occupational carcinogen thought to act by an epigenetic mechanism.

### **Current Controversies on Carcinogenesis Testing**

One of the most significant debates in modern toxicology unfolded in response to several papers, particularly three published in late 1990 in the *Proceedings of the National Academy of Sciences*, by Dr. Bruce Ames, pioneering microbial biochemist and the inventor of the in vitro “Ames assay” for mutagenic activity. These articles, published back-to-back, expressed his concern that too many chemicals are testing positive in animal bioassays and that industrial chemicals are regarded as responsible for cancer even though people ingest a much greater number and variety of “natural” carcinogens in food. In this debate, the originator of the most significant alternative to the animal bioassay—a test that has led to the identification of numerous suspect carcinogens—cast doubt on the validity of the “gold standard” of toxicology, against which his own assay is compared. The controversy this stimulated has great implications for occupational medicine and even greater implications for environmental and consumer health regulations.

The protocols used in animal carcinogenesis bioassays involve prior, routine toxicological testing to identify the “maximum tolerated dose” (MTD), followed by scaled dosage at several points to extrapolate a curve. From this basic model, and from whatever epidemiological information is available, low-dose extrapolations are made. Unfortunately, animal bioassays for carcinogenicity are very expensive; most cost more than \$1 million if established protocols with good laboratory practice are used. Ames suggests that when an animal is exposed to the MTD, toxicity at the cellular level results in necrosis or severe cell injury followed by cellular proliferation to repair the damage. The increasing rate of mitosis and tissue growth is then more

susceptible to carcinogenicity as a result of erroneous DNA repair, production of genotoxic oxidative compounds as a result of inflammation (as suggested above for asbestos and foreign-body epigenetic carcinogenesis), and errors in gene duplication. This is not a new idea. It was discussed decades ago (in less mechanistic terms) as the most plausible mechanism for “scar carcinoma” (emergence of an adenocarcinoma at the site of an old parenchymal scar in the lung, usually a granuloma from tuberculosis) and is under active discussion today as a possible mechanism for cancer associated with exposure to silica or asbestos (see section on epigenetic mechanisms of cancer, above).

Ames cited as evidence for this idea a rough correlation between toxicity and carcinogenic activity and the exposure-response curve, which is linear-quadratic rather than linear for many carcinogens. Ames proposes that if cell proliferation occurs, the increased cell turnover as toxicity is achieved will increase the likelihood of a mutagenic event. This has important implications for setting an occupational and environmental exposure standard. If Ames is correct, there are likely to be thresholds for carcinogenesis and therefore safe exposure levels. This would provide a scientific basis for setting allowable exposure standards for carcinogens with a presumption of safety, rather than assuming that any exposure leads to some finite risk of cancer. (Nonlinearity is also explained at least as well by the saturation of host defense mechanisms, however.)

Ames is also concerned that “three wrong assumptions” have been made by investigators who conduct and interpret animal bioassays. He lists these assumptions as:

1. Only a small proportion of chemicals are likely to be carcinogens. (This assumption, however, underlies regulatory policy, not science.)
2. Testing at high dosage would not lead to carcinogenic effects unique to the high dose. (Contrary to the impression Ames gives, this issue has been under active discussion in toxicology circles for years; the rationale for high levels of exposure are concerned with shortening latency periods and overcoming

- host defenses, not mimicking effects at low levels. As a practical matter, rodent bioassays could not be done otherwise.)
3. Chemical carcinogenesis would be explained by the mutagenic potential of chemicals. (This is a much more recent idea that was largely promulgated by Ames himself. It has been accepted for many years that there are other mechanisms of carcinogenesis besides direct interaction with DNA; benzene, for example.)

Ames suggests that synthetic chemicals are no more carcinogenic than “natural” pesticides present in foodstuffs and presumably consumed by humans over evolutionary time. His argument for this basically consists of the repetition of lists of plausible plant pesticides present in significant amounts in the human diet that are among the many chemicals secreted by plants normally or when they are injured or spoiled as a form of self-protection against pests. He argues that this problem may have been exacerbated by the development of more pest-resistant agriculture. He does not explain how this is plausible, however, given that the latencies are not compatible and that spoilage was undoubtedly a much bigger problem in the past, before modern storage technology. None of the chemicals cited by Ames as matching or exceeding the toxic potential of industrial chemicals have been established, either individually or collectively, as significant causes of cancer in developed communities, with the possible exception of animal fat. These plant products in foodstuffs are probably far older than he suggests, although some pest-resistant plants have been bred in recent years by maximizing production of these chemicals. However, human beings have very likely evolved effective mechanisms for dealing with the most common plant pesticides, because there is no reason to think that they are recent plant adaptations and every reason to think that levels of ingestion because of food spoilage were much higher in the past. His discussion also begs the question of the many “anti-carcinogens” that are also present in foods.

Ames’s line of argument explicitly stems from his impression that too many rodent carcinogens have been identified in testing programs to date to be plausible as likely risks to health. However, those

chemicals tested first in programs such as the National Toxicology Program were largely selected on the basis of some reason for concern. Therefore, one might expect the earliest results to be largely positive. Only recently have scientists gotten around to testing chemicals that are not under some suspicion and that are therefore more likely to test negative. In Ames's view, the "correct analysis" is to compare the proportion of rodent carcinogens among a random group of industrial chemicals versus a random group of natural chemicals. He does not explain why such a comparison might be meaningful, or for what purpose this procedure would be correct.

Toxicologists are well aware of the limitations of animal bioassays. The methodological issues are under active investigation in many laboratories, such as the National Center for Toxicological Research, in an effort to find new ways of accounting for interspecies differences and structure-activity relationships, and for interpreting pathological end points other than cancer. Ames calls for further research but does not acknowledge that the scientific community was working on these issues long before he became aware of them.

The track record of animal bioassays has actually been very good; the assays tend to be sensitive but not specific. The techniques of animal bioassays have become more refined with time, and there is indeed reason to believe that increasingly inbred strains of rodents are becoming more susceptible to cancer, both spontaneous and induced. However, many demonstrated animal carcinogens, such as vinyl chloride, have later been shown to be human carcinogens. Only one known human carcinogen—arsenic—has not been shown to be carcinogenic in animals and the explanation for this, which involves a difference in metabolic pathway, is known.

By contrast, Ames's own assay—and a number of other *in vitro* testing procedures—has been less useful than was originally thought. This does not diminish Ames's earlier achievement; it only points out the difficulty of the problem.

There are indeed data that suggest a close correlation between toxicity and carcinogenicity; this has been cited in support of the cell proliferation hypothesis. However, there are also many exceptions to

this generalization. Many toxic chemicals exert their action without widespread cell necrosis (and at doses far lower than the MTD), and the site for carcinogenesis is not always the same as it is for primary toxicity. An effect of cell proliferation may well be to condition tissue to increase the rate of mutagenic events, but that is consistent with the intention of the animal bioassay. It may be a mechanism, not a cause, and may well underlie the carcinogenic effect of asbestos and silica, among other nontraditional carcinogens.

There is one statement of Ames with which many can agree: "What is chiefly needed is to take seriously the control of the major hazards that have been reliably identified, without diverting attention from the major causes by a succession of highly publicized scares about factors that may well be of little or no importance as cause of human disease." Ames is an eminent scientist who has become aware of the complexity of the problem of environmental chemical exposure. In his capacity as a member of a California state technical committee to advise on listing chemicals as putative carcinogens, he has been regularly faced with the knotty problem of interpreting the toxicological literature. In these articles, he is challenging a public policy that is based on a low threshold for concluding that a chemical is a human carcinogen. If there is a social reason to rethink action based on scientific evidence, that issue should be addressed on its own merits. It is not clear that attacking the scientific basis for the animal bioassay for cancer is a responsible way to change public policy.

## **REPRODUCTIVE AND DEVELOPMENTAL TOXICOLOGY**

Modern society is well aware of environmental and occupational exposures and the risk to the next generation. This concern is one of the most powerful factors in the perception of risk by the public (see Chapter 7). "Reproductive toxicology" addresses issues of interference with reproduction, with the principle adverse outcome being infertility. Reproductive failure may be the result of a specific toxic effect on reproductive organs or toxicity to the

mother interfering with carrying the child to term. “Developmental toxicology” addresses issues of embryonic and fetal development, therefore assuming that the fetus has survived and implying toxicity in a different, lower range with more-specific effects. After birth, infants and children continue to develop and may be affected by exposures in their environment, including in the school and at home.

### **Reproductive Toxicology**

Reproductive toxicology became a major issue in 1977, when a certain pesticide used to kill nematodes (roundworms), dibromochloropropane (DBCP), was found to have contaminated the water supply of a pesticide plant in California, resulting in infertility in the male workers. DBCP was found to act on the Sertoli cells by a unique mechanism. The DBCP case study is still the clearest example of a clinically significant reproductive effect from an occupational exposure.

Notwithstanding human population dynamics, reproduction in the human organism is very inefficient compared to most species. It is true that human beings, unlike other mammalian species, do not experience episodic reproductive receptivity (“heat” or “rut”) and therefore experience few restrictions on time of mating; they also do not inhibit fertility as a function of population density, as do other species. However, for the most part reproduction in human populations is slower and more prone to failure than in other mammalian species. Sperm production is less intense; the number of eggs is fixed; the time to sexual maturity is relatively long (at thirteen years or so); spontaneous abortion is very common; females reach an effective end to fertility at menopause; and mating requires elaborate behaviors and is controlled by relatively rigid cultural norms. It does not take much to interfere with normal reproduction and to disrupt the process. This can result from toxic effects on the parents, endocrine disruption, genotoxicity (interfering with normal gene expression), disruption of implantation in the uterine wall, embryonic or fetal

toxicity, teratology (developmental defects that are not viable), and disruption of sexually related behavior (especially libido).

In general, the male reproductive tract is much more susceptible to interference from toxic exposures or hormonal factors than the female reproductive tract, although the female tract is much more complicated. This susceptibility to disruption is high in utero—before birth—and drops off in childhood, returning during puberty. Anything that interferes with cell division has the potential to suppress fertility in both sexes, but particularly in men.

The basic human body plan is female, and that is how the body develops unless a gene on the Y chromosome induces hormonal factors that suppress development of the female reproductive (“Müllerian”) tract, thereby allowing the male (“Wolffian”) tract to develop. Fertility (and the onset of puberty) in women can be affected by environmental factors such as heat, diet, drugs, chronic disease, and behavior, but especially by age.

There are a limited number of adverse reproductive outcomes that can be observed:

- Infertility
- Spontaneous abortion (often undetected and appearing as menstrual irregularity)
- Stillbirth (recognized spontaneous abortion)
- Congenital defects (developmental outcome)
- Low birthweight
- Maturation failure
- Delayed adverse effects (such as neurocognitive impairment or cancer risk)

In men, the principal adverse outcomes are infertility, low sperm count, abnormal sperm morphology, reduced sperm motility, altered sexual performance (erectile dysfunction or loss of libido), and gynecomastia (resulting from hormone disruption). The target tissues

of concern are in the testes: cells on the sequence to become gametes (spermatogonia, spermatocytes, spermatids), cells that support them (Sertoli cells), endocrine tissue (Leydig cells), and accessory structures (the epididymis, for example). Representative exposures that affect the male reproductive system include ethanol, pesticides, PCBs, TCDD, PAHs, metals (especially lead and mercury), solvents, and tobacco smoke.

For women, the principal adverse outcomes are menstrual irregularity (which may represent implantation failure, through the second week), adverse reproductive outcomes, the risk of breast cancer (possibly related to hormone disruption), and endometriosis. Although occasionally a cause of distress, nausea and olfactory hyperacuity (heightened sense of smell) are not abnormal during pregnancy. The potential target organs are much more varied: the hypothalamus-pituitary axis in the central nervous system for both the menstrual cycle and for ovulation, the ovary, the ova itself, the uterus and especially the endometrium, and the uterine tract itself. Representative exposures that affect the female reproductive system include PCBs, PBBs (polybrominated biphenyls, withdrawn from commerce), formaldehyde, metals (arsenic, lead, and mercury), pesticides, phthalates (which make polymers pliable), and solvents.

Careful studies of reproductive outcomes are consistent in showing that about 12 percent of couples in North America are infertile and that at least 20 percent of pregnancies result in spontaneous abortion. Of those pregnancies that end in live births, 7 percent are low birthweight ( $<2500$  g), and 18 percent are preterm ( $\leq 35$  weeks). These figures demonstrate that successful reproduction is not a foregone conclusion.

### **Endocrine Disruption**

Endocrine disrupters have the potential to interfere with hormonal control of reproduction. Xenobiotics in water have been shown to have an endocrine-disrupting effect on other species, particularly

fish. A highly publicized example is the ambiguous genitalia found on male alligators in Lake Apopka, Florida, which is heavily contaminated because of, among other things, a direct spill of the pesticide dicofol into the lake in 1980. Organochlorines (including PCBs and various pesticides), nonyl phenol, and phthalates have the potential to mimic or inhibit hormone activity, which in the case of the non-coplanar PCBs may include thyroid hormone activity. Steroid hormones, phytoestrogens (plant-derived, short-acting, estrogen-like substances), and diethylstilbestrol may exert a direct estrogenic effect.

There is no question that exposure to some chemicals may induce hormonal abnormalities in human beings. The major question with respect to public policy is whether this effect occurs in human populations as a result of environmental exposure. The relevant examples involve high levels of exposures in the workplace or taking pharmaceuticals. An incident in Puerto Rico involving an outbreak of gynecomastia among workers involved in the manufacturing of oral contraceptives showed that the potential exists for human effects. From the 1940s through the 1960s, a synthetic estrogen mimic, diethylstilbestrol (DES), was used as a treatment to prevent high-risk pregnancies (to the extent that it became the standard of practice), although it was ineffective. DES induced clear cell adenocarcinoma of the vagina in some daughters of mothers who were given the drug, and it is said to have induced genital tract abnormalities in many of the daughters and some sons. The potential is clearly there for endocrine mimicry, but these examples involve much higher concentrations than are observed in environmental exposures.

Definitive evidence for a human effect at environmentally relevant concentrations has been elusive. Perhaps the strongest evidence for this has been the demonstration (since replicated elsewhere) that children in the Great Lakes region born to mothers who consumed more than an average amount of locally caught fish (but not extreme amounts) were born with lower neurobehavioral test scores than their less-exposed peers, and that this group difference continued through childhood with reduced academic performance. This is attributed to exposure to PCBs possibly interfering with thyroid

hormone activity early in development of the central nervous system (CNS).

Other, more-heavily publicized evidence has not been as persuasive. There was an apparent decline in the sperm count and sperm quality over the decades of the twentieth century, but the evidence is not conclusive. In part this is because sperm counts are rarely done on people without suspicion of a problem, so there is a strong selection bias. There also appears to be a difference between urban and rural counts, both historically and in contemporary studies. Increasing reports of hypospadias and endometriosis would appear to suggest a problem, but exposure data are weak (especially with the withdrawal of many candidate chemicals, including pesticides and consumer products, from commerce over the years). The evidence, though controversial, suggests that the incidence and mortality of breast cancer does not seem to bear a relationship to environmental endocrine disruptors.

A particularly difficult outcome to interpret is altered sex ratio in populations. Many epidemiological studies have now been performed to determine the ratio of male to female births (normally 51 to 49, with male children having a lower survival rate in infancy) and have detected a tilt toward more female births. The effect has been observed in many populations exposed to pesticides, air pollution, and radiation, but has also been reported in the aftermath of floods and other disasters and among mothers with diabetes during pregnancy as well as eclampsia and preeclampsia, which occur long after the sex of the offspring has been determined. It is therefore unclear whether this outcome reflects a biological effect.

## **Developmental Toxicology**

Birth defects are common but are perceived by the public as rare. This is in part because minor birth defects often pass unnoticed except to a concerned parent, and in part because of the emotional impact of major and obvious birth defects. In fact, malformations of

the heart and great vessels occur in 1 in 115 live births; club foot, 1 in 735; cleft lip or palate, 1 in 930; spina bifida, 1 in 2,000; for a total of approximately 2.5 percent of live births. The rate of birth defects in miscarriages and stillborn births is much higher, reflecting the loss ("fetal wastage" is the term of art in reproductive biology) of fetuses that are not viable.

When a birth defect occurs, it is entirely understandable that parents wish to know why it happened. Chemical exposures become the focus of attention. The cause of most birth defects is not known, which leaves ample room for speculation and anxiety. The unusual cases in which chemical exposures have played a role in outbreaks of birth defects tend to reinforce the belief that many birth defects are caused by environmental or occupational exposures, when in fact such outbreaks have been rare.

Historically, the study of birth defects has been called "teratology," after the Greek word for monster, and chemicals or physical exposures that cause birth defects are called "teratogens." Indeed, throughout history birth defects have been perceived as monstrous, and many folk theories have developed to explain them, often on supernatural grounds. One of the more lasting is that psychological shocks or frightening events that occur to the mother affect the fetus. In fact, 43 percent of birth defects have no known cause; 28 percent appear to be the result of mutation; 23 percent appear to be multifactorial in origin; and 3 percent are attributed to uterine insufficiency and twinning (which carries a high risk). Current thinking in reproductive biology is that about 3 percent of birth defects are attributable to known reproductive hazards, including alcohol, tobacco smoke, and drug abuse.

Teratogenesis can only follow low to moderate toxicity because high toxicity would lead to infertility, fetal loss, or maternal death. The chemicals that do this, therefore, are usually otherwise well tolerated in the body. Unlike reproductive toxicology, which is mostly about the male reproductive system, developmental toxicology is almost entirely about exposure of the mother as the pathway of exposure to the fetus, rather than toxicity also involving the mother.

Maternal toxicity may also affect the fetus, but the effects are non-specific and usually not associated with specific birth defects.

Only about forty chemicals, most of them drugs, have been shown to produce prenatal developmental abnormalities in human beings. Others have been demonstrated in animal models, but most teratogens show a high degree of specificity by species. For example, cleft palate (a failure of facial structures to meet and fuse during development) is induced by cortisone in rabbits and mice but not rats. Susceptibility of the fetus to a teratogen depends on gene-environment interactions and varies by developmental stage of the fetus. Along the way, there are many influences on the fetus, with various effects depending on the stage of development. These include the health of the mother, infection (especially rubella, cytomegalovirus, toxoplasmosis, and syphilis), diabetes, hypoxia, alcohol intake, drug abuse, tobacco smoking, radiation, heat or hyperthermia, and diet (especially insufficient folate).

A wide range of xenobiotics are associated with developmental abnormalities. Many of them reflect high exposure levels and are consumed in relatively large quantities. Tobacco smoke and nicotine alone are associated with developmental abnormalities, which include low lung function, even when exposure occurs before birth. Ethanol is associated with fetal alcohol syndrome. Drugs of abuse, especially cocaine and heroine, have characteristic syndromes as well as withdrawal effects. Retinol and vitamin A are potent teratogens affecting the neural plate, which is why great care is taken to prevent exposure in women of child-bearing age when retinoic acids are used for the treatment of acne. Occupational and environmental exposures include metals (lead and mercury), ethylene oxide, toluene (a potent neurotoxin), and d-penicillamine (a drug that is sometimes used for chelation in metal toxicity).

Birth defects may result in fetal death (which is expressed as infertility if early in embryonic development and as stillbirth or miscarriage in the fetus), viable anatomic malformations, growth retardation (either low birthweight or delayed maturation), functional deficits (such as neurocognitive impairment), and behavioral

abnormalities (probably including attention deficit/hyperactivity disorder). Functional deficits and behavioral abnormalities may be associated with prenatal exposure to metals, particularly lead and mercury (see chapter 10).

Anatomic malformations occur when interference occurs with normal development during a critical phase of organogenesis. The blastocyst, which becomes the developing embryo, forms four to six days after conception in the human being (three to five days in the rat, which is the usual experimental model), and implantation follows by a day; embryotoxicity up to this point will probably not be perceived. Early toxicity, within the first two weeks, will cause loss, but because it occurs so early it will usually be perceived, if at all, as menstrual irregularity rather than a spontaneous abortion. Organogenesis begins on day twenty-one and proceeds to day fifty-six, with the neural plate forming between day eighteen and day twenty. Interference with development during the critical period of the individual organ will cause a malformation in that organ. Major malformations typically occur before week nine because the developing tissues become less susceptible to fetotoxicity. The heart develops from weeks three to nine; genitalia, weeks seven through infancy; and the CNS, weeks three through infancy and childhood. Limb buds, which will form the arms and legs, appear around day twenty-nine or thirty, and the limbs develop through weeks four to ten.

The major example of developmental toxicity that illustrates the specificity of fetotoxicity with respect to organogenesis is thalidomide. Thalidomide is a hypnotic, analgesic, antiemetic, and anti-inflammatory drug currently used in the treatment of leprosy and HIV/AIDS. It was first synthesized in 1956 and was tested on animals, showing no detectable toxicity. Thalidomide was marketed in the 1960s in Europe as an exceptionally safe analgesic because the drug had few side effects, except for peripheral neuropathy, which developed only with long-term use. The manufacturer filed for approval with the U.S. Food and Drug Administration as well, but a skeptical staff scientist, Frances Kelsey, held up approval of the drug because she felt strongly that the information supplied was

incomplete and that it minimized the potential significance of the neuritis. In 1961 a hospital in Hamburg reported a dramatic leap in frequency of a rare birth defect involving limb reduction (truncated limbs with shortened or missing long bones) called “phocomelia,” and thalidomide was found to be a consistent risk factor. Ultimately, 5,850 infants were born malformed with phocomelia worldwide due to thalidomide, but few cases occurred in the United States, and those almost entirely among Americans who had obtained the drug in Europe. Other birth defects included congenital heart disease and eye, intestinal, and renal malformations. Elaboration of the mechanism was very difficult because human beings, almost uniquely, turned out to be the most sensitive species. Thalidomide has a short half-life, so the teratogenic effect was very restricted, occurring only with intake between days twenty and twenty-six of pregnancy.

Another instructive example is Bendectin, a popular antiemetic used for severe nausea during pregnancy (“morning sickness”) in the 1970s. The evidence is clear that Bendectin was not, in fact, associated with reproductive or developmental toxicity. However, the drug became the issue in an important legal action, which reached the Supreme Court of the United States: *Daubert v. Merrell-Dow*, in which it was alleged that Bendectin caused birth defects. Issues of the admissibility of scientific evidence led to the formulation of the Daubert decision, a key decision. Although the decision was for the defense, and the alleged toxicity of Bendectin was effectively refuted during the trial, many other lawsuits followed and the manufacturer withdrew the drug from the market because of the cost of defending it.

## REGULATORY TOXICOLOGY AND RISK ASSESSMENT

The application of toxicology to risk assessment is often called “predictive toxicology,” and the application to risk management is often called “regulatory toxicology.” Since the 1970s, the increasingly formal

approaches of risk assessment and “risk management” (see Chapter 7) have guided the development of toxicological methods in response to the practical data needs of regulatory and health protection.

Toxicology is essential to chemical “hazard identification,” the first stage of the risk assessment process. Usually, toxicological research or testing results in the recognition that exposure at a certain level may be hazardous and suggests outcomes or mechanisms that may be important in human exposure. Toxicology defines exposure-response relationships in animal studies, target organ systems (the most likely clinical outcome for human toxicity), and possible susceptibility states. These aspects of the toxicity of the agent are critical in interpreting the likely response in humans at lower exposure levels and projecting the possible impact of exposure in populations (as discussed in Chapter 7).

Problems in toxicology are divided conceptually between the assessment of effects of short-term, high-level exposures and those of long-term, low-level exposures. The first is usually a clinical problem with low levels of uncertainty, and the second is usually a regulatory problem with high levels of uncertainty. Predictive and regulatory toxicology and risk assessment deal most often with the latter case.

Toxicological testing begins with short-term, high-level “acute” exposures. Acute toxicity is seen in human cases of uncontrolled occupational situations, suicides, and accidents, and animal studies of acute toxicity are useful in dealing with such situations. However, acute experimental toxicity studies are valuable in providing an indication of the potential toxicity of the agent, identifying the organ systems involved, providing data to support extrapolation to lower concentrations, and identifying mechanisms that may be involved in host defense or the expression of injury. Because cancer latency periods tend to be shortened after high exposure levels, short-term, high-level exposures to carcinogens are also valuable in providing a practical means for identifying the potential for carcinogenesis of an agent within the life span of a group of animals (and the investigator’s grant).

The National Toxicology Program was established by the U.S. government specifically to address issues involved in extrapolating to long-term, low-level exposure situations. Improved models of dosage and cumulative exposure, characterization of species and strain differences, identification of basic biological mechanisms, and description of a few intensively studied models of exposures carried to exceptionally low levels have laid a foundation for regulatory toxicology. (Because of power considerations, as described in Chapter 3, these studies require huge numbers of subject animals and are colloquially called “mega-rat studies” by many toxicologists.) Such studies are much too expensive to be conducted on every agent of interest.

### **Experimental Design**

Selection of any animal model should take into account its biological relevance to the application for which it is intended. For studies to assess that degree of hazard to humans, the experiment must consider human routes of exposure, metabolic pathways, and the potential for expression of the effect. The animal model must duplicate these characteristics or at least their most important features.

Strain differences within species are as important as species differences. Inbreeding has resulted in considerable differences among rat strains in their response to longer-term exposures. There is evidence that in some strains of rats, the background or spontaneous frequency of cancers has risen greatly, which may complicate interpretation if comparisons are made with earlier generations. The age and sex of the animals may also be important. Young animals may be more resistant to toxic effects than older ones; for example, neonatal mice are more resistant to oxidant gases than are older mice. In other cases, young animals are more susceptible, which is probably more often the case with human children. In some instances, as in studying the effects of aromatic hydrocarbons that may mimic the effects of estrogens, female animals may be more susceptible than males.

The life span of the animals dictates what studies are practical. In long-term studies, it may be important to allow an animal to live out

its entire span rather than to sacrifice it at the end of a specified period. For example, in early studies on cancer in rats exposed to silica, the animals were sacrificed after two years and did not develop tumors; animals allowed to survive as long as possible did develop lung tumors. With the general population living much longer than in the past, long-term survival studies in animal models are growing increasingly important for other outcomes in which age may be a contributing factor. For example, it is now apparent that some neurological conditions induced by neurotoxic chemical exposures may be silent until loss of neurons due to aging results in a clinical expression of dementia or movement disorders (in the case of Parkinsonism).

Following exposure, the time course of the experiment will depend on whether the investigator is concerned with acute or chronic effects, the timing of the recovery process, sequelae of the acute effects, or late effects that follow a latency period.

The outcomes that may be studied in an experiment in toxicology are limited only by imagination, methodology, and available apparatus, but there are some constants. Mortality and changes in weight and appearance are always recorded. When experiments provide unexpected results or seem anomalous, these data are often useful in sorting out what happened and as indicators of acute toxicity in chronic experiments. Morphologic changes, reflected either in grossly visible changes or histological changes, are the most common outcome reported otherwise. Biochemical and immune changes can often be monitored serially by drawing blood during the course of the experiment. Behavioral toxicology uses as outcomes animal behavior in standardized situations that can be directly observed, such as how animals respond to a novel problem or changes in conditioned behavior.

The hazard presented by exposure to a toxic agent is usually summarized by three benchmarks: median mortality estimates, the minimum exposure at which adverse effects are detected, and the maximum exposure at which no adverse effects are observed. A hybrid terminology born of toxicology and epidemiology has developed for applications in risk management, as shown in Table 2.4.

**Table 2.4.** Terms in Predictive Toxicology

*Acceptable Daily Intake (ADI).* The amount of a xenobiotic that can be taken into the body over a prolonged period of time without appreciable health risk over a lifetime. In the United States, ADIs are primarily used for food additives; in Canada they are often used more broadly in assessing risk against a benchmark.

*Acute test.* Toxicity tests that evaluate effects immediately or shortly after exposure.

*Chronic.* Toxicity testing that lasts longer than or toxic effects that become visible more than 90 days after the initiation of the experiment.

*Draize Test.* Usually, the Draize Eye Irritancy Test, a standard method since 1944 of testing the potential of chemicals to cause eye irritation. Mostly applied to toiletries and cosmetics, the test involves instilling a diluted solution of the chemical into one eye of a rabbit and then comparing effects with the contralateral, untreated eye. Toxicologists rarely use this test today, primarily because of concern for animal welfare but also because species differences make rabbits less susceptible than humans to eye irritation. The test was primarily used to confirm the safety of ingredients in cosmetics and for chemicals for which there was no prior indication of toxicity. It was developed as a test after an incident in which an adulterated mascara product caused blindness in women who used it.

*LC<sub>50</sub>.* Median lethal concentration. That concentration of the agent in air that is estimated from empirical evidence to result in death of just one-half of the subject animals exposed; because this term refers to a concentration rather than a dose, a time period for the exposure must be specified. Used for gases and particles. (Could be used in aquatic toxicology to refer to water concentrations, but extrapolation to humans would be limited.)

*LD<sub>50</sub>.* Median lethal dose. That amount of the agent administered that is estimated from empirical evidence to result in death of just one-half of the subject animals exposed; refers to single-dose or cumulative administration.

*LOEL.* Lowest observable (adverse) effect level. The lowest exposure observed to result in any toxic effect in subject animals. The U.S. Environmental Protection Agency uses the abbreviation LOAEL, for “no observable adverse effect level.” In conventional predictive toxicology, any functional change is usually considered adverse to the organism unless it is nutritional.

(Continued)

**Table 2.4.** Terms in Predictive Toxicology (*Continued*)

---

**MOE.** Margin of exposure. The difference between an ADI or an RfD and the corresponding NOEL for the exposure. The larger the MOE, the smaller the risk is presumed to be, regardless of other uncertainties.

**MTD.** Maximum tolerable dose. The highest dose that an animal can tolerate for prolonged periods without observable toxic effects, except for carcinogenicity. Determines the highest dose given in carcinogenicity studies.

**NOEL.** No observable effect level. The highest exposure level tested that has not resulted in any toxic effect in the subject animal. The U.S. Environmental Protection Agency uses the abbreviation NOAEL, for “no observable adverse effect level.” In conventional predictive toxicology, any functional change is usually considered adverse to the organism unless it is nutritional.

**RfD.** Reference dose. By definition of the Environmental Protection Agency, a dose taken orally that is presumed to be acceptably safe, usually by reference to a NOAEL. RfDs were originally used as part of the risk management framework for pesticides but have come to be used more widely as a point of departure, for standards setting. The Agency for Toxic Substances and Disease Registry recognizes RfDs as applying to all exposure pathways.

**Skin test.** A test of skin irritancy as opposed to carcinogenesis. Usually, the Draize Skin Test, in which the chemical is applied in solution to the shaved skin of a rabbit; rats and mice are sometimes used, especially when the outcome of interest is cancer rather than irritation. The endpoint is skin irritation. John H. Draize, a toxicologist at the Food and Drug Administration, invented both this test and the eye test that now bears his name.

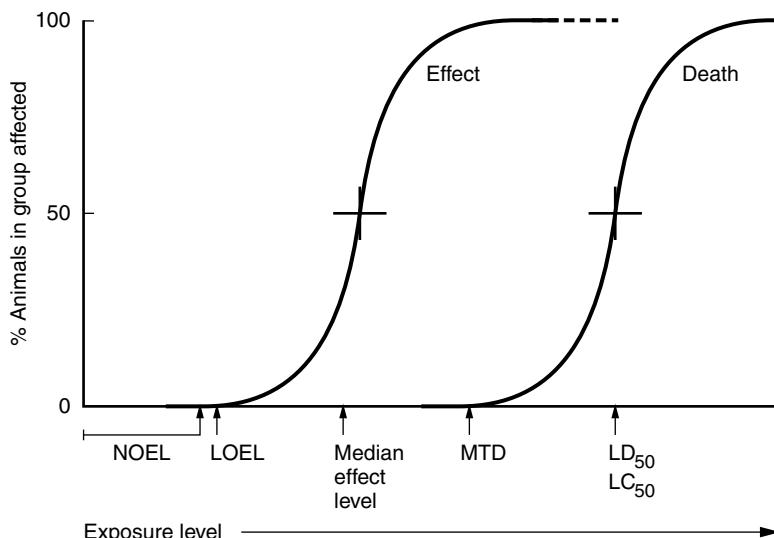
**Subchronic test.** Toxicity tests that evaluate test subjects 90 days after initiation of exposure, long enough to permit recovery from initial, acute effects but not long enough for chronic outcomes, such as cancer, to be observed.

**Threshold.** The level of exposure or dose that is associated with the earliest effect or the first appearance of an outcome. In practice, the threshold is almost never known. However, it is bracketed by the LOEL and NOEL for the particular effect being studied.

---

The “lethal dose” and “lethal concentration,” at which 50 percent of animals do not survive exposure, are abbreviated LD<sub>50</sub> and LC<sub>50</sub>. These two measures of median lethality are key terms that are remarkably reproducible and permit comparison of the relative potency of agents in a single species and the relative susceptibility of several species to a particular agent. LD<sub>50</sub> and LC<sub>50</sub> are derived by calculation from experiments in which populations of animals are exposed to the agent at different levels (usually four). Population-based, or “epidemiologic,” exposure-response curves are plotted, and the exposure level estimated to be associated with 50 percent mortality is determined by a statistical technique called “probit analysis.” The animals are subjected to necropsy and the pathologic findings noted in detail. Although the LC<sub>50</sub> and LD<sub>50</sub> data are limited to mortality, these studies still provide a great deal of basic information at relatively low cost.

By contrast, LOEL, “the lowest observable effect level,” and NOEL, the “no observable effect level,” are less conceptually concrete. These apparent threshold levels are less reproducible in experimental studies because they are more sensitive to experimental design and conditions than the LD<sub>50</sub> or LC<sub>50</sub>. New methods of measurement in laboratory studies may uncover previously unsuspected effects, resulting in constant re-evaluation of the LOEL and NOEL. The LOEL is sometimes assumed to be the toxicological threshold of response for the xenobiotic, but it is actually the first point above the threshold that happens to have been sought and identified. The NOEL is sometimes treated as an estimate of the threshold (especially in risk management), but it is only the highest level measured without detectable effect using then-current methodology. Over time, as increasingly sophisticated experiments are performed and more subtle effects are recognized, reported LOELs are getting lower. One cannot know with certainty whether one is far from the true threshold or very close until a set of paired, close-together NOEL and LOEL thresholds can be observed, preferably in the same experiment or at least by consistent methods. These important benchmarks are illustrated in Figure 2.33. The NOEL is more reliably a “no effect” level the further away it is from



**Figure 2.33.** Benchmarks in regulatory toxicology.

the LOEL; however, the further away, the less useful it is in defining the maximum tolerable exposure for risk assessment purposes.

The traditional approach to setting exposure standards was to identify the NOEL and to reduce it by a safety factor of 10 (when good exposure data for humans are available), 100 (when data are relatively weak), or 1,000 (when few useful data are available). This traditional and robust approach to standards setting has been replaced in large part by identifying a probably safe level (often the NOEL) and reducing it by a series of uncertainty factors that happen to be 10 or close to it. In other words, the process is not much different at all. What is different is that more is known about the NOEL than in the past, and safety factors are less arbitrarily applied. Uncertainty or safety factors applied for purposes of regulating occupational exposure standards are less than those for environmental exposure standards, based on assumptions regarding the vulnerability and susceptibility of the population at risk of exposure. The working population exposed is considered more likely to be healthy and unlikely to contain highly susceptible individuals.

### *Toxicity Testing*

In vitro tests are usually used as “tier one” tests: rapid and inexpensive screens for problem compounds, as the first level in a series of tests to define the toxicity profile of an agent, or as part of an investigation to determine the mechanism of toxicity of an agent. They cannot be used alone for toxicity evaluation. With increasing pressure from animal rights activists to reduce the use of animals in biomedical research and product evaluation, there has been an increasing effort to develop more reliable and versatile in vitro tests, but they are not yet ready to replace in vivo test systems.

Application of these general approaches to specific problems varies with the intended use of the chemical. The most stringent requirements for toxicity testing, as one might expect, are in the pharmaceutical industry. Pesticides and agrichemicals are also subject to extensive testing, as are cosmetics and food additives. Until recently, systematic testing was uncommon on new products destined for industrial applications; as a result, the data base for evaluating occupational exposures has been weak. With recent legislative developments recognizing the workers’ right to know the effects of exposure to chemicals and the rise of product and third-party liability litigation, testing of chemicals before introduction in industry is becoming more common. The requirements for testing are weakest for chemical exposures that may occur in the environment, except for pesticides. If the chemical is not a pesticide, and is not otherwise regulated, there are no specific requirements for testing in the United States or Canada before the agent can be released into the environment. Handling the discharge would be a matter of applying existing pollution-control laws. In Europe, this issue is being addressed by new chemical safety regulations.

The extremely high cost of lifetime studies using rodents makes it impractical to test every compound of potential interest. Recently, Lave et al. have proposed that these expensive assays be limited to agents of greatest social impact, as determined by a complex mathematical function that takes into account the social cost of a false

positive or a false negative result, and either the results of first-tier testing or the likelihood of the agent being carcinogenic on the basis of known structure-activity relationships for the class of chemicals. This ingenious proposal makes intensive demands on an uncertain testing technology and the unproven theories of chemistry, but adoption of the approach, at least implicitly, is almost inevitable because all candidate compounds cannot possibly be tested given available resources. Improved understanding of the biology of these testing procedures and the basis for extrapolation between species will almost certainly make this approach more viable in the future.

A major issue in selecting any kind of animal model is the biological relevance of the model to the application intended. The experiment must be at least comparable to human routes of exposure, metabolic pathways (if applicable), and the potential for expression of the effect. Strain differences within species are as important as species differences. Inbreeding has resulted in considerable differences among rat strains in response to longer-term effects. The longevity of animal species places constraints on what can be studied. Animals that survive less than two years in confinement, such as mice, are difficult to use for long-term exposure studies. Rats do survive this long, but full expression of the effects of exposure may require the animal to live out its life span rather than be sacrificed after an arbitrary time period. The age and sex of the animals are also important considerations, as noted above. The toxicity of a new chemical or an agent suspected to be toxic is usually assessed by a sequence of studies, each level of which is called a “tier.” A “tier one” study, for example, may involve screening laboratory tests that do not involve animals—collectively called *in vitro* studies—such as the Ames assay (described below), to identify potential carcinogens early and to exclude them from further consideration as a possible industrial product. A “tier two” study might involve determination of LD<sub>50</sub> or LC<sub>50</sub>. Higher tiers may involve “subchronic” studies (ninety-day exposures, followed by sacrifice of the animals to examine sublethal effects); “chronic” studies of six months or a year; lifetime studies (to evaluate carcinogenicity over two to three years); and special studies to examine developmental or

reproductive effects, toxicokinetics and metabolism, allergenicity, phototoxicity, and behavioral effects. As the tests become more sophisticated and regulatory bodies require outcomes that are more difficult to detect, product testing has become much more expensive.

Each test in a tier has its own significance, depending on the anticipated end use of the agent and its potential toxicity. Each tier builds on the information gained by the previous tests, in principle. For example, the first test to be performed in a feeding situation may be a “range-finding” test to determine the exposure levels associated with lethality in small groups of animals, perhaps using exponentially scaled dosages such as 1, 3, 10, and only three animals per group (3 is close to the square root of 10, so this point is useful in calculating the LD<sub>50</sub> on a log plot). Based on the exposure level that resulted in mortality of all or most of the animals, the next experiment can be scaled accordingly to include exposure levels at four points, some of which are likely to fall above and below the LD<sub>50</sub> and allow its estimation with reasonable accuracy. Once the LD<sub>50</sub> is known, exposure levels well below the LD<sub>50</sub> can be studied to determine sublethal effects.

Subsequent toxicity testing varies in the United States depending on the use of the chemical proposed, the amount that is expected to be released, and the regulatory structure governing the class of chemical. Pharmaceuticals are subjected to rigorous testing. Pesticides are subjected to testing levels almost as rigorous, including tests of eco-toxicity (the effect on the environment). Industrial chemicals and consumer products are rarely tested in comparable depth and may not be tested at all if the constituent chemicals are already deemed to be safe, the quantities are small, or an exemption is granted by the Environmental Protection Agency.

In 2001 the European Union introduced the REACH program (Registration, Evaluation, and Authorization of CHemicals), which requires much more stringent testing of chemicals in commerce, both novel chemicals and those already in use. The regulations and programming to support REACH have now started, and REACH is beginning to affect international trade by establishing a new,

evidence-based standard manufacturers must meet in order to export their products to world markets. REACH is much more data-intensive than existing U.S. regulations. This development will require a considerable increase in toxicity testing worldwide.

Increasing use is made in toxicity screening of in vitro testing, structure-function relationships, and modeling in order to reduce the use of animals in biomedical research and product evaluations.

A major issue in toxicity testing is Good Laboratory Practice. Most products other than pharmaceuticals are evaluated under contract by commercial laboratories, using standardized protocols. Several highly publicized incidents came to light in the 1970s involving falsified data, poor quality control, inadequate study design, and the use of nonvalidated methods. A uniform code of Good Laboratory Practice was then devised by the Food and Drug Administration to qualify acceptable studies for regulatory purposes, and a similar GLP code was developed by the European Union and the Organization for Economic Cooperation and Development. These codes are now followed by all reputable laboratories.

### ***Bioassays for Carcinogenicity***

Carcinogenesis bioassays are essential tools for health protection against carcinogens. They are protocols that are used to determine the risk of cancer through experimentation in intact subject animals, almost always rats. The challenge of cancer bioassays is to expose the animals to a level that is carcinogenic but not highly toxic, and then to support the animals until the end of their (usually) natural lives. Cancer bioassays used to be routinely truncated at two years (which is close to the natural lifespan of a rat), but it was found that some late-appearing cancers had been missed. Groups may be sacrificed early in order to document the time of appearance of tumors, or at different dosage levels.

Cancer bioassays generally involve a range-finding experiment that identifies the “maximum tolerable dose (MTD),” the maximum dose or level of concentration that does not kill the animal. The

MTD will be lower than the LD<sub>50</sub>, of course, and may be thought of as the highest NOEL that can be found for lethality. The MTD is then used as the highest dose or exposure level for a long-term study, usually at several dosage levels, conducted with controls. Controls are important because, as noted, several rodent strains have shown increasing rates of background tumors over the years, and because occasionally environmental factors (such as the laboratory chow) have a modifying effect on cancer risk. Historical controls are therefore unreliable. Following sacrifice or death of the animals, a necropsy is performed to identify cancer in specific organs, following a strict protocol to be sure that nothing is missed.

As in the case of LD<sub>50</sub> studies, statistical considerations related to exposure and response apply to the study groups the same as if they were small populations in epidemiological studies (see Chapter 3). Because the animal populations are small, their power is very low at low levels of exposure to carcinogens. This means that the yield of cancers and the probability of detecting an effect are low if the bioassay is performed at levels encountered in most occupational and all environmental situations. It is therefore necessary to push the level of exposure as high as possible, first, in order for the animals to initiate a sufficient number of cancers to be statistically different from the background rate among control animals (particularly at sites where rats commonly develop tumors, such as lung); second, to yield enough tumors to study histologically for variations; and third, to shorten the latency period so that they will be developed within the animals' life span. Cancer bioassay studies are often criticized for using unrealistically high exposure levels, but this is the rationale and reason why it cannot be otherwise.

### ***In Vitro Testing for Mutagenicity***

The drawback to animal bioassays is that they are very expensive; it is impossible to do one now for less than \$1 million using established protocols with good laboratory practice. For this reason, scientists have searched for accurate but inexpensive alternative screening tests. This is the origin of interest in in vitro assays.

In 1975 the same Bruce Ames introduced the salmonella revertant bioassay for mutagenic activity, with and without activation by mammalian hepatic enzymes. This was considered such a breakthrough that for a time it appeared that animal bioassays could eventually be dispensed with altogether. The best-known in vitro test, the Ames assay is an elegant test for mutagenicity that detects interaction with DNA by induction of a back-mutation in bacteria from a nutrient-dependent strain to the wild type, as indicated by growth on a nutrient-deficient medium. A positive test in any of several strains, each with a sensitivity profile different for various chemical classes, correlates strongly but not absolutely with carcinogenicity in animal species.

Many other in vitro tests have been and are being developed, so laboratories today have a battery of such tests, almost all for the purpose of identifying potentially genotoxic agents. Relatively few practical in vitro assays have been developed that use mammalian cells.

The problem with the Ames assay and the other in vitro bioassays that were developed later is that they detect mutagenic (genotoxic) activity in a life form that is different in important respects from mammalian cells: cell-wall permeability, chromosome packaging, intracellular enzyme activity and location, and absent host-defense mechanisms, to name a few.

Mutagenicity in the Ames assay does suggest a high probability that the same chemical will cause cancer in animals, but there are some exceptions. There are more exceptions the other way: a negative response in the Ames assay is not always predictive that a chemical will be free of carcinogenic activity in animals or humans. Many known carcinogens act by “epigenetic” means rather than directly on the genome. This means that they cannot be detected by the Ames assay or by other conventional in vitro genotoxicity assays.

After its limitations came to be recognized, the “Ames assay” assumed a more modest place as a “first-tier” test that, if positive, indicates an increased probability that a chemical will show carcinogenic activity in an animal bioassay, and is therefore a warning sign.

A negative test, particularly for metals, hormones, and carcinogens suspected of acting by epigenetic means, is not considered to be definitive evidence of an absence of mutagenicity; but it is reassuring, and the chemical continues in the screening protocol. As valuable as in vitro testing may be, it cannot replicate the complexities of the intact animal. As a result, in vitro tests are usually used as “tier one” tests, rapid and inexpensive screens for problem compounds already in use, nonspecific indicators of exposure to an excretable carcinogen by testing urine, or part of an investigation to determine the mechanism of toxicity of an agent.

## BIOLOGICAL MONITORING AND TRACKING

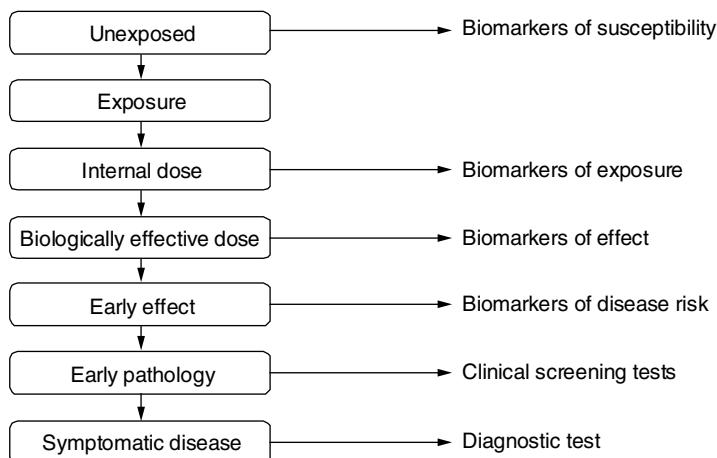
One approach to worker protection is to introduce individualized tests that measure exposure in the individual rather than clinical impairment or abnormality. By monitoring xenobiotic levels, their metabolic products, or their physiological effects in the body, a truer picture can be pieced together of exposure from all sources; of the risk for a particular individual, subtracting out the variations in the workplace or environment; and of trends over time in that one individual. These methods are designed to quantify levels of internal dose following exposures lower than would be implicated as a cause of illness. They are not the same as toxicological screening tests, which assume exposures already within the toxic range.

“Biological monitoring” is a current term for the periodic evaluation of workers by detecting indicators of exposure through trace residues of the agent, its metabolites, or a biological response. This may take the form of measurements in body fluids or expired air. End-expired breath carbon monoxide levels give a very precise indication of exposure to carbon monoxide during a shift. Likewise, end-expired breath solvents measured by mass spectroscopy give a very accurate picture of the individual exposure to these agents, at levels far below the current threshold limit values (TLVs). In its original use, the term meant monitoring by biological mechanisms of adaptation or changes

in function, not by determining xenobiotic levels directly. For example, red blood cell cholinesterase levels are often used as an indicator of biological effect suggesting exposure to organophosphate pesticides; plasma cholinesterase levels are more sensitive but volatile and are more often used clinically. The language has changed over thirty years, and biological monitoring is now used to embrace toxicological testing, although the author prefers to keep the distinction and to use the term “biomonitoring” for the more general sense of surveillance by testing.

Biomarkers are measurements that can be used to assess states, risk factors, or characteristics in a subject without invasive procedures or clinical diagnosis. Biomarkers should not be confused with clinical screening tests. Sensitivity, specificity, and predictive value do not apply to biomarkers, because they are not clinical tests, at least in this application. There may be an association with an outcome, but it is usually remote. (Figure 2.34 illustrates the various types of biomarkers.)

Biomarkers and biomonitoring methods are used for exposure assessment in epidemiological studies (see Chapter 4), monitoring individual worker exposure (by the American Conference of Governmental Industrial Hygienists (ACGIH)) as “biological exposure



**Figure 2.34.** Biomarkers and their relationship to exposure and response.

indices” (BEIs), forensic analysis, tracking individual exposures over time to gain knowledge regarding environmental exposures, and drug testing (a very specific application that is discussed in Chapter 19).

The National Research Council recognizes three types of biomarkers, each with different interpretations related to exposure and response (see Chapter 1), to which may be added, for clinical and surveillance purposes, a fourth for disease risk:

- Biomarkers of susceptibility
- Biomarkers of exposure
- Biomarkers of effect (or response)
- Biomarkers of disease risk

A “biomarker of susceptibility” is an indicator that an individual is predisposed to a given condition or to a particular response to a xenobiotic exposure. The NRC defines it as an indicator of an inherent or acquired limitation of an organism’s ability to respond to the challenge of exposure to a specific xenobiotic substance. This is not exactly the same as a genetic marker, which identifies a particular genotype or risk for genetic disease, such as  $\alpha_1$ -antiprotease deficiency. The biomarker of susceptibility reflects a phenotype such as a metabolic pathway, a greater chance of multifactorial disease risk, or a relative deficiency in a corrective mechanism (such as DNA repair).

A “biomarker of exposure” is formally defined by the NRC as an exogenous substance, its metabolite, or the product of an interaction between a xenobiotic agent and some target molecule or cell that is measured in a compartment within an organism. Basically, a biomarker of exposure can be a structural or biochemical change—such as adducts—that suggests exposure to a xenobiotic, but the term is also used for detection of xenobiotics and their metabolites in body fluids. For example, carboxyhemoglobin is a biomarker of exposure to carbon monoxide and can be monitored in blood or air, unchanged except for its affinity for hemoglobin. The ACGIH BEI for the blood level is set at 3.5 percent of hemoglobin, and for

the air level (which reflects blood levels closely) at 20 ppm in expired air. Another example is cadmium, which will also be used to illustrate a biomarker of effect. Cadmium can be monitored in blood (5 µg/l) or urine (5 µg/g creatinine). Some xenobiotics are measured unchanged, such as acetone in expired air or urine, and in other cases metabolites are used, such as the xylenes (there are three), which are metabolized to methylhippuric acid and measured in urine.

A “biomarker of effect” is formally defined by the NRC as a measurable biochemical, physiological, or other alteration within an organism that, depending on magnitude, can be recognized as an established or potential health impairment or disease. To summarize, a biomarker of effect is basically an indication of a response to a xenobiotic; this response may be adaptive or at an early stage of toxicity. The implication is that the effect measured by the biomarker is somewhere on the pathway to a toxic response. When a biomarker is an early indication of toxicity or disease, it may also be considered a screening test for secondary prevention, but biomarkers by design are supposed to identify changes that take place much earlier than the disorder would express itself (see Chapter 8). An example of a biomarker is cadmium, which can be monitored directly in urine or by an effect: the presence of  $\beta_2$ -microglobulin in urine (which indicates irreversible change and is therefore a late indicator). Another example is the serum cholinesterase level, which is an indirect measurement of exposure to organophosphate pesticides. Aniline produces methemoglobin in blood, which can be expressed as a percentage of total hemoglobin.

A “biomarker of disease risk” (not an NRC term) is an indication that a person is at risk of developing an overt disease and may already be in the early stages, such as PSA for prostate cancer. A biomarker of disease risk may be applied to acquired disease as well as susceptibility states and to a large extent simply represents a renaming of familiar clinical screening tests. (Biomarkers of disease risk are discussed more fully in Chapter 8.)

A direct biomarker is collected in a “matrix,” or medium, which may be expired air, blood, serum, urine, hair, or saliva. The matrix

should be easy to collect by noninvasive means and not prohibitively painful to obtain (venipuncture is problematical for routine measurement for workplace BEIs, but acceptable for occasional monitoring, for example for environmental tracking). It should be easy and socially acceptable to work with (ruling out feces, earwax, and nail clippings). It should not inconvenience the subject or worker (making 24-hour urine collections impractical for workplace monitoring). The matrix should be easily dissolved and practical for laboratory analysis (another reason why cerumen, nail clippings, and feces are not suitable). The xenobiotic or its metabolite is extracted from the matrix and concentrated if necessary. Testing will have the usual detection level and can be interpreted against a reference range, which is not the same as a set of “normal values” in other laboratory tests. Unlike normal values in a laboratory test, there is no “normal range” because, except for a very few elements that are also nutrients, such as iron, or essential trace elements, such as magnesium, the body does not exert homeostatic or adaptive control to keep the levels in blood roughly constant. The body does not treat solvents as it does sodium or albumin, in other words. The reference range for biomonitoring usually reflects a geometric distribution, not a normal distribution, with the mode or expected level in unexposed workers close to zero. A major advantage to BEIs is that virtually all workers should be protected at this level.

Biomarkers cannot be used when the xenobiotic or its metabolite is also a product of normal metabolism at relatively high concentrations. The body produces carbon monoxide endogenously from bile metabolism, so there is some baseline value of carboxyhemoglobin for all people. If it were higher, carboxyhemoglobin or expired-air carbon monoxide could not be used. BEIs are not helpful when the adverse effects are local or superficial, such as skin or mucosal irritation or bronchial inflammation. They are not useful with respect to allergy, because sensitization thresholds are not related to what they measure.

“Tracking” is biomonitoring applied to populations for the purpose of assessing environmental exposures and determining trends. The U.S. Centers for Disease Control and Prevention call it

Environmental Public Health Tracking. Several U.S. federal agencies and some states (notably, California) have developed programs for following levels of exposure in their populations to see if they are rising or falling and to determine geographic patterns. Eventually, it is expected that associations between disease and environmental exposures will be uncovered. This requires the collection of huge amounts of data and maintenance of extensive databases, necessarily with individual identifiers for longitudinal studies, and a formidable task of interpretation. Unlike exposure assessment for other purposes (see Chapter 4), tracking is not undertaken for purposes of compliance or to test a hypothesis for research. The data collected by tracking can be used for research, and biomonitoring and tracking protocols are built into programs such as the National Health and Nutrition Survey, which pioneered the effort, and the National Children's Study in an effort to determine future disease risk.

## RESOURCES

- Greenberg M, Hamilton R, Phillips S, McCluskey GJ. *Occupational, Industrial, and Environmental Toxicology*. St Louis: Mosby; 2003.
- Hoffman RS, Nelson LS, Howland MA, Lewin NA, Flomenbaum NE, Goldfrank LE. *Goldfrank's Manual of Toxicological Emergencies*. New York: McGraw-Hill; 2007.
- Klaasen C, ed. *Casarett and Doull's Toxicology: The Basic Science of Poisons*. 5th ed. New York: McGraw-Hill; 2007. [The essential textbook of toxicology.]
- Reigart JR, Roberts JR. *Recognition and Management of Pesticide Poisonings*. Washington, DC: US Environmental Protection Agency, EPA 735-R-98-003; 1999. [Updated and Web accessible at [www.epa.gov/pesticides/safety/healthcare](http://www.epa.gov/pesticides/safety/healthcare).]

## NOTEWORTHY READINGS

- Ames BN, Gold LS. Chemical carcinogenesis: too many rodent carcinogens. *Proc Natl Acad Sci USA*. 1990;87:7772–7776.
- Ames BN, Magaw R, Gold LS. Ranking possible carcinogenic hazards. *Science*. 1987;236:271–280.
- Anderson D. *In vitro models*. *Drug Safety*. 1990;5(suppl 1):27–39.
- Ball P. *The Devil's Doctor: Paracelsus and the World of Renaissance Magic and Science*. New York: Farrar, Straus, and Giroux; 2008.

## The Praeger Handbook of Occupational and Environmental Medicine

- Clewell HJ III, Andersen ME. Use of physiologically based pharmacologic modeling to investigate individual versus population risk. *Toxicology*. 1996;111:315–329.
- Doul J. The past, present, and future of toxicology. *Pharmacol Rev*. 1984; 36:155–185. [Classic.]
- Fung VA, Barrett JC, Huff J. The carcinogenesis bioassay in perspective: application in identifying human cancer hazards. *Environ Health Perspect*. 1995; 103(7):680–683.
- Greenberg MI, Hamilton RJ, Phillips SD, McCluskey GJ, eds. *Occupational, Industrial, and Environmental Toxicology*. 2nd ed. St Louis, MO: Mosby; 2003.
- Guidotti TL. Exposure to hazard and individual risk: when occupational medicine gets personal. *J Occup Med*. 1988;30:570–577.
- Guidotti TL. Hydrogen sulphide. *Occup Med*. 1996;46:367–371.
- Guidotti TL. Acute cyanide poisoning in prehospital care: new challenge, new tools for intervention. *Prehosp Disaster Med* 2006;21(2)(suppl):S40–S48.
- Guidotti TL, McNamara J, Moses MS. The interpretation of trace element analysis in body fluids. *Indian J Med Res*. 2008;128:524–532.
- Klaasen CD, ed. *Casarett & Doull's Toxicology*. 6th ed. New York: McGraw-Hill; 2001.
- Lauwerys RR, Hoet R. *Industrial Chemical Exposure: Guidelines of Biological Monitoring*. 3rd ed. Albany, GA: Lewis, 2001.
- Lave LB, Ennever FK, Rosenkranz HS, Omenn GS. Information value of the rodent bioassay. *Nature*. 1988;336:631–633.
- Maibach HI, Feldman RJ, Milby TW, Serat WF. Regional variation in percutaneous penetration in man. *Arch Environ Contam Toxicol*. 1971;23:208–211.
- Plaa G. Animal models. *Drug Safety*. 1990;5(suppl 1):40–45.
- Richards IS. *Principles and Practice of Toxicology in Public Health*. Sudbury, MA: Jones and Bartlett; 2008.
- Ward JM. The two-year rodent carcinogenesis bioassay—will it survive? *J Toxicol Pathol*. 2007;20(1):13–19.
- Witschi H. Some notes on the history of Haber's Law. *Toxicol Sci*. 1999;50:164–168.

# 3 EPIDEMIOLOGY

Epidemiology is the study of the distribution of disorders, states of health or well-being, and potential determinants of disease risk in populations. Its essential concern is to identify the factors influencing the distribution of disease risk, but it also represents the best opportunity in occupational and environmental medicine (OEM) to study disease in human beings, because experimental methods are limited. Epidemiological investigations provide the knowledge base regarding human health that undergirds occupational and environmental protection, evidence-based clinical practice, and much of risk analysis. Epidemiology provides a systematic approach to monitoring the health experience of populations and identifying associations between exposure and outcome. Together with toxicology, epidemiology is a basic science of OEM.

Epidemiological research provides the essential knowledge base for evidence-based practice, causation analysis, regulatory policy, risk assessment, health promotion, and the identification of workplace hazards. It is an unusual week or month that an occupational and environmental physician with any degree of responsibility does not encounter a problem requiring interpretation or evaluation of an

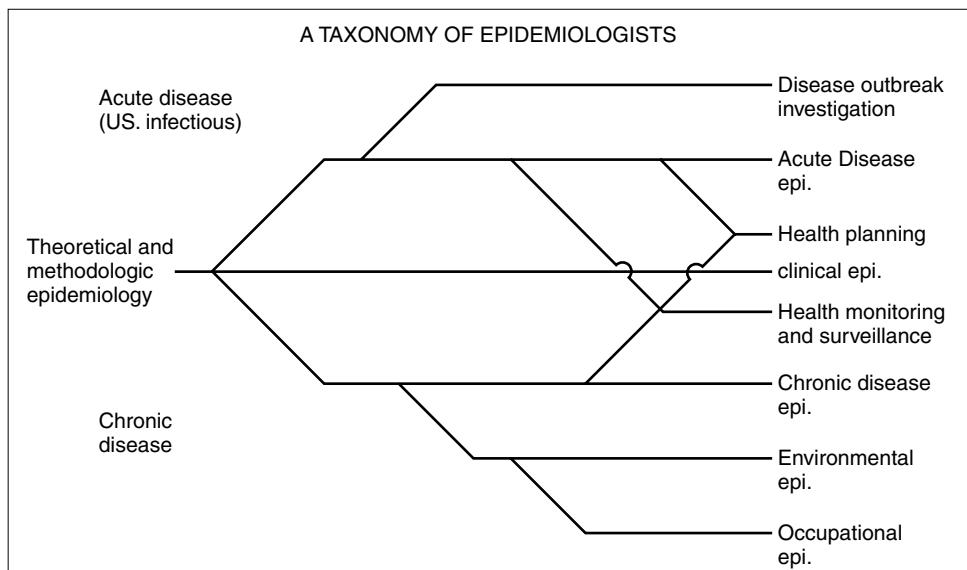
epidemiological study. A working knowledge of epidemiology is therefore essential preparation for an OEM physician.

This chapter is not a primer on epidemiology and does not try to cover the field to the depth of a standard textbook in epidemiology. Rather, it is written to help the OEM physician with limited exposure to epidemiology to interpret studies. Rather than emphasizing methodology, this chapter describes why and how we use epidemiology in etiological investigations and basic study designs, and limitations of epidemiology that are relevant to OEM. For this reason, it is written “backward” compared to a textbook of epidemiology. It starts with interpretation and inference and then proceeds to the basic study designs and study characteristics that need to be examined, and only then examines data sources. Other material is included to assist the OEM physician in determining when a study might be feasible and assessing a proposal to conduct a study.

Historically, epidemiology began as a means of investigating patterns of disease or death frequency for clues to a cause and as the search for public health interventions that would control outbreaks. Most of the early subjects of epidemiologic investigation were infectious diseases, such as cholera, tuberculosis, and malaria, but even in the earliest, formative years of epidemiology as a discipline, attention was paid to chronic disorders of noninfectious causes, such as nutritional deficiencies. Throughout its development as a science—from its earliest beginnings in the examination of crude British vital statistics data at the municipality level—there has been a strong theoretical and methodological emphasis, as well as a strong practical emphasis on public health applications in epidemiology. For a time, some epidemiologists (most notably Olli Miettinen) saw their field as a basic, abstract science addressing frequencies and phenomenology as clues to underlying causes, with applications well beyond health. Most epidemiologists, however, see the science as a very practical, disciplined way of observing connections and making inferences about cause and effect in public health. Over time, epidemiologists have specialized, and the historically broad division between “acute disease epidemiology” (mostly but not exclusively infectious disease)

and “chronic disease epidemiology” (mostly but not exclusively diseases of aging and cancer) has become more nuanced and less distinct. The OEM physician is mostly concerned with occupational and environmental epidemiology, which are considered sister branches of chronic disease epidemiology, although both fields are also interested in diseases that might be considered acute or that have their origins in infectious agents. Figure 3.1 illustrates the major themes in epidemiology and the interrelation of various schools of thought in the field.

Epidemiology has pursued a pragmatic tradition of making the best possible use of data collected for other purposes and using incidents of exposure and illness outbreaks as models for more fundamental insights. It has also developed techniques and approaches that do not rely on absolute accuracy. In fact, one of the most important attributes of epidemiology as a science is that it is unusually tolerant of inaccuracy and robust in dealing with faulty data. The erroneous perception



**Figure 3.1.** A taxonomy of epidemiology, showing the historical division between “acute” and “chronic” disease, which has become less relevant over time.

that epidemiological studies are always invalidated by the limitations on the data on which they are based has been a major stumbling block in the acceptance of epidemiological findings by physicians, which is why it is mentioned at the outset.

Some question the validity of epidemiological studies on the grounds that the data on which they are based is not precise or accurate (the difference between the two is explained in Chapter 4). Death certificates, in particular, have been essential to occupational and environmental epidemiology, and their accuracy has sometimes been a limiting factor in study efficiency. Physicians know that death certificates are often filled out by busy clinicians without direct knowledge of the case, and that hospital discharge summaries or clinic visits are not always coded correctly. However, accuracy in data sources matters much less than is often assumed. This is not a situation of “garbage in, garbage out.” As long as the error or imprecision in a data set is reasonably consistent within the population under study and is the same in magnitude, pattern, and frequency for the external reference group (those to whom the subjects are compared), all that is lost is efficiency. Statistics and the “law of large numbers” (as population size increases, patterns are reliably described by probability) work strongly in favor of epidemiologists. It simply takes more subjects (sometimes many more) to detect the effect, or “signal,” against the background of “noise” introduced by misclassification.

Epidemiologists often refer to themselves as “acute” or “chronic” disease epidemiologists, depending on whether they study acutely infectious disease or chronic conditions such as cancer and cardiovascular disease. In fact, this simple categorization is no longer relevant. It does not begin to reflect the diversity of the field. It does not recognize that some infectious diseases are chronic or have become so through better treatment options, such as HIV/AIDS, and that some environmental and occupational outcomes are acute. It also fails to convey the increasing trend toward interdisciplinary studies involving laboratory and epidemiology methods. Even so, “chronic disease epidemiology” is shorthand in the discipline for a set of methods and approaches that are best suited and often developed for

etiological studies on the effects of environmental and occupational exposures. These methods were largely developed in the latter half of the twentieth century.

When epidemiology underwent its expansion after the Second World War, most epidemiologists were physicians. Physicians had the great advantage of understanding in depth the nature of the disease they were studying, which was helpful in study design and field work. However, relatively few physicians (although those few were exceptional and highly influential) had the rigorous statistical and mathematical training or insight required to be highly innovative or to get the most knowledge out of data analysis. Few physicians took graduate study in epidemiology, except for what was incorporated in the master's in public health (MPH) degree, which quickly became insufficient. As the discipline matured, it attracted many more students from outside medicine. At the same time, physicians became increasingly scarce in medical research generally, because of unfavorable economics and increasing difficulty in balancing clinical medicine with a rigorous research career. By the 1970s, epidemiology was increasingly dominated by PhD epidemiologists working in research and by master's degree-level epidemiologists in public health practice. As a broad generalization, that generation of non-medically trained epidemiologists took the field to new heights of sophistication and did studies that were much better technically, but they often did not have the same depth of understanding of the problem as their physician predecessors.

The pendulum swung again in the late 1980s and 1990s when the field attracted many investigators with backgrounds in molecular biology and genetics, drawn in large part because of the new emphasis on biomarkers and gene-environment interactions. The methods that are available for what used to be called "molecular epidemiology" or "genetic epidemiology," and is now often called "genomic epidemiology," can be very accurate and sophisticated, but the limiting factor tends to be the accuracy of exposure assessment, which is much less than the accuracy of the outcome measures. This is a critical problem because the key issue in modern environmental and

occupational epidemiology is the interaction between genetic contribution (what used to be called “constitutional factors”) in the individual and exposure to environmental hazards, such as chemical agents. As important as it is, genetic factors tend to drive disease susceptibility in relatively few individuals; the majority of people are not highly susceptible to most environmental exposures, and the contribution of genetic predisposition is not great. It is therefore of only limited value to have a precise estimate of the contribution of genetic factors to disease outcome when the contribution is small in the population and not decisive in the individual, and the uncertainty introduced by inaccuracy in the assessment of exposure is orders of magnitude greater. The current emphasis on genomics may discover important examples that point in the opposite direction, but for the most part it is not clear that genetic makeup is the most important factor in the expression of disease for most environmental hazards or for most people.

Today, epidemiological research reflects more biological sophistication than ever, but faces a new challenge. Large population studies have become increasingly difficult because of restrictions on access to data, often motivated by concerns for privacy, and cutbacks in funding, especially for large-scale prospective studies in occupational health (defined below). Epidemiologists hitting their stride in the field today are often well-grounded in biomedical sciences, even if they are not physicians, but have learned to make better use of data sets originally compiled for other purposes, and often have to settle for methods and approaches that are robust and cost effective rather than ideal for investigation of the problem. Thus, in a sense epidemiology is coming full circle from its origins, again using as a primary data source vital statistics compiled for administrative purposes.

The introduction of improved methods in epidemiology led to a considerable surge in the number and quality of studies in environmental and occupational health after about 1975. In turn, much of the progress achieved by the science of epidemiology over the last several decades has been driven by the need to resolve methodological problems in environmental and occupational epidemiology, especially

with respect to difficult issues arising out of research on asbestos, air pollution, and occupational cancer.

In parallel and at about the same time, the subdiscipline known as “clinical epidemiology” was being refined as the essential tool of evidence-based clinical medicine and applied to treatments and clinical outcomes. Although much of clinical epidemiology cannot be directly applied to occupational medicine, because clinical trials are rarely feasible in the field, the methodology and the body of evidence-based medicine that has accumulated over the years are increasingly being applied to clinical management in occupational injury care, workers’ compensation, and disability prevention. This connection will be explored further in the chapters on clinical OEM.

The type of epidemiological investigation of greatest practical concern to OEM is that usually termed “occupational epidemiology,” which is primarily devoted to etiological research and has the advantage of studying populations of workers, who are typically exposed to the agent of interest at significant levels. Occupational epidemiology is closely related in subject and methodology to “environmental epidemiology” but tends to concentrate on more tightly defined population groups and relatively higher and more consistent exposure levels. Occupational and environmental epidemiology both build on the individual, or toxicological exposure-response relationship (as discussed in Chapter 2), to establish an association and to identify opportunities for intervention that may control problems encountered in the workplace. Environmental epidemiology tends to examine exposures on a larger scale and at lower concentrations. Its findings are applied more often to public health and regulatory policy for exposure of the general public, and usually do not have direct application clinically or in the analysis of causation (see below and Chapter 23). The two subfields are closely related and use most of the same basic study designs, but differ in their application, the characteristics of the populations under study, data access, and many technical issues of implementation and methodology. For example, there is no counterpart to the important “healthy worker effect” (defined later) in environmental epidemiology, but it is critically

important in occupational epidemiology. On the other hand, the “ecological” study design (described later) is almost never used in occupational epidemiology but is common (and badly abused) in environmental epidemiology. Where there are differences, they will be treated as separate fields of epidemiology. Where no distinction is made, the discussion refers to both.

Epidemiology provides OEM not only with its pertinent practical knowledge but also a framework—a coherent thought process—that emphasizes patterns and behavior in populations rather than in individuals. It provides powerful tools and even more powerful definitions and conceptual approaches in addressing issues in OEM. For occupational medicine specifically, epidemiology provides essential methods for tracking and assessing management of hazards in groups of workers. Of greatest value, however, is that epidemiology provides data on exposure and response directly pertinent to human beings. Epidemiology therefore provides an opportunity to confirm, extend, and interpret findings that may have been initially observed in animals or that can later be addressed in laboratory models. It also provides the opportunity to identify and observe the characteristics of health effects in human populations in the absence of a relevant experimental system.

Epidemiology’s great strength is that it is a science of generalizations, so it does not depend on the exceptions and idiosyncrasies of the individual. However, epidemiology’s great weakness is that it is a science of generalizations, and therefore cannot be applied uncritically to the problems of the individual. It can only describe patterns that may or may not—are likely or unlikely to—apply to a given individual. For the occupational and environmental physician, this makes epidemiology an invaluable tool for surveillance of workplace or environmental hazards, addressing the needs of the workforce or community, or investigating occupational or environmental problems, but a very blunt instrument when applied to assessing risks for an individual or the causes of a disorder in the individual case. In such situations, epidemiology is limited to providing valuable guidance and a most likely estimate of future risk or past causation

(assuming that the individual conforms to the characteristics of the population or subgroup to which he or she is compared).

## **CAUSATION**

The concept of causation is critical to understanding epidemiology and interpreting its findings. Much of occupational and environmental epidemiology is devoted to finding and documenting the causes of health outcomes and is guided by knowledge gained from epidemiology. “Causation,” as used in North America, is the assessment of the probable cause of a disorder or condition in the individual case. “Causality” (used in Britain as a synonym for “causation”) properly means the process of cause and effect itself—a distinction important mostly in philosophical discussions. Causation is a key issue in designing effective prevention, tort litigation, and forensics, as well as adjudicating complicated workers’ compensation claims. Assumptions about causation also underlie regulatory standards-setting and surveillance. These applications depend heavily on epidemiological evidence from studies directed at etiology, the cause of disease.

### **Etiological Investigation in Epidemiology**

Most epidemiological studies that apply to OEM are directed at etiology, although clinical epidemiological studies of outcomes and best practices have become more popular in recent years. Etiological studies in this field are not easy, and even excellent studies have many methodological and practical limitations. For this reason, occupational and environmental epidemiology are at the center of an ongoing debate regarding the value of epidemiology and how informative the discipline really is. This is largely a sterile debate, fuelled in part by efforts to discredit inconvenient findings and in part by the backlash against “junk science,” but at its core is a philosophical misunderstanding regarding the nature of scientific investigation.

The problem of causation is inextricably bound to the issues of “proof” and what epidemiology can and cannot do. Because

occupational and environmental epidemiology rarely involve controlled experiments, inferences must be made on multiple observations and treated statistically in order to discern patterns. Investigators in other sciences, such as experimental toxicology (but not clinical toxicology), have the advantage of being able to conduct controlled experiments. Epidemiology comes close to controlled experimental methodology in clinical trials, but these are generally not feasible otherwise and so are very rare in the field of OEM. Epidemiology is therefore often described as predominantly an “observational” rather than experimental science, not unlike astronomy, particle physics, or, classically, the social sciences. Each of these sciences must infer from the data what probably happened and look for inconsistencies (“anomalies”) in the data to prove their theories wrong, because they cannot structure an experiment to test the theory directly. Epidemiological data in occupational and environmental health are therefore indirect and uncertain as evidence for cause and effect. Some critics of epidemiology have therefore taken the position that the field has little value and cannot be trusted. This is an exaggeration.

Like all sciences, epidemiology cannot directly prove anything to be true. No amount of data or compilation of evidence can definitively prove any assertion in science, although working assumptions are accepted as to which theory is probably true. Science—whether an experiment in the laboratory or a study using epidemiological methods—can only demonstrate that something is false. Science advances by testing hypotheses—reasonable predictions from theory—to prune away those that are wrong, and then infers from the evidence what is probably correct, or at least tenable, as a working theory. This process, called “falsification” in the literature of the philosophy of science, is no different for epidemiology than for any other science. Epidemiology is simply a means of testing theories in science by using population-based data.

No science can prove anything to be true. Not really, although strong and converging evidence can leave little remaining doubt. No amount of data or compilation of evidence can definitively prove any

assertion in science, because proof implies certainty in all respects and no possibility of an alternative explanation.

“Falsifiability” is an essential characteristic of scientific theories. The leading philosopher of science, Karl Popper, used the illustration of an investigation of the color of swans: one may observe thousands of swans over many years and conclude reasonably that all swans are white. However, this assertion cannot be proven by observation alone, and the observation of a single black swan falsifies and therefore invalidates it. Although it may be more difficult to falsify in epidemiology than in other fields, because of statistical uncertainty and the observational nature of the discipline applied to occupational and environmental health, the working theory can often be ruled out. In other words, a well-designed and well-powered epidemiological study is like any other scientific observation: it can invalidate a hypothesis or provide empirical observation that can be used to infer a general principle.

### **Epidemiological Evidence for Causation**

Causation is a key issue in tort litigation, forensics, and adjudicating complicated workers’ compensation claims. Assumptions about causation also underlie regulatory standards-setting.

The association of lung cancer with cigarette smoking is indisputably causal and provides a clear example. The way it happens in real life is that smoking increases the probability that a smoker will get lung cancer but not every smoker will get it, and some nonsmokers, although very few, do get lung cancer regardless. A few cases of lung cancer might be missed in the counting (if, for example, a patient were to die of a cancer-related cause and no autopsy was performed); a few might be present but undetected when the subject died of unrelated causes; and a few cases might be misdiagnosed as cancer when they represented some other pathology, although this is unlikely. Cigarette smoking, as a risk factor, also goes along with other risk factors, such as drinking alcohol, and it interacts with certain exposures on the job that also cause lung cancer. These days,

older people are more likely to smoke than younger people, and it is known that as people age, they are also more likely to get lung cancer; therefore age has to be taken into account and “adjusted” in the analysis. The job of an epidemiologist interested in cigarette smoking and lung cancer is to sort it all out and to discover the association between the outcome (lung cancer) and the exposure of greatest interest (cigarette smoking) despite everything else that is going on in the causation of lung cancer. Imagine, however, that a naive epidemiologist studying lung cancer and objects in the home did not realize that cigarette smoking was important and instead mistakenly counted the number of ashtrays in a person’s home. That epidemiologist might falsely conclude that ashtrays cause lung cancer if he or she went only by the statistical evidence of an association. Although this may seem to be an obvious example when cigarette smoking is used, one may imagine the potential for such errors when studying populations exposed to multiple airborne hazards (including such formidable hazards as asbestos), air pollution, and cigarette smoking taken all together, simultaneous as they are in real life.

### **The Hill Criteria for Assessing Likelihood of Causation**

The most widely accepted set of criteria for assessing the likelihood of causation reflected in epidemiological data is that proposed by Sir Austin Bradford Hill, a British biostatistician of distinction who introduced the criteria in 1965 and later popularized them in a seminal textbook of statistics applied to medicine. He was searching for guidelines that suggested that an observed association truly represented cause and effect rather than something else. The criteria he devised are presented in Table 3.1. The criteria are intended to be applied rigorously and as a group; the more that appear to be satisfied, the more likely it is that the association observed is truly causal.

The Hill criteria for accepting an association as causal are not equally robust, however. Some criteria are clearly stronger than others. The strength of an association, for example, is a strong criterion. The standardized mortality ratio (SMR) is an estimate of risk in

**Table 3.1.** Sir Austin Bradford Hill's Criteria for the Plausibility of a Causal Association

- 
1. Strength of the association
  2. Consistency among studies, especially by different techniques
  3. Specificity of outcome
  4. Exposure precedes disease outcome
  5. Dose-response relationship (epidemiological)
  6. Plausibility of a biological mechanism
  7. Coherence of chain of evidence
  8. Experimental association, especially dose response (toxicological)
  9. Analogy to similar effect produced by a similar agent
- 

which 100 represents the risk of the general population (or other reference). SMRs less than 150, for example, are usually considered unlikely to be strongly associated with a single work-related exposure. A strong association may be statistically uncertain, with a wide confidence interval around the estimate, particularly if the number of cases is small. This is a common problem with rare disorders, as illustrated by the association between phenoxy herbicide exposure and non-Hodgkin's lymphoma or soft-tissue sarcomas.

The criterion for consistency presents a paradox. Contradictory results from a study using different methods do not directly refute the original observation, because the circumstances of the investigation may not be easily comparable. Discrepancies among studies often provide greater insights than consistencies. In epidemiology, many of the differences in findings between similar studies have to do with differences in methodology and the populations under study, because even in occupations with great similarity from place to place, such as firefighters, no two studies are identical. Results from an identical study conducted on a different population sample, which is possible only in theory in epidemiology, would help in the evaluation of consistency but nothing else. They would contribute no new insights, unless major anomalies were observed, and serve only as a replication for purposes of statistical inference testing. Few investigators in epidemiology try to precisely duplicate studies

conducted by others, and funding sources are not interested in supporting them, so they are not available. Therefore, as a practical matter, this Hill criterion has more to do with methodological consistency, generalizability, and convergence of findings than true consistency in outcomes.

Specificity is the weakest criterion of all. It is now well established that one exposure (for example asbestos) may lead to a number of outcomes (asbestosis, bronchogenic carcinoma, mesothelioma, and probably carcinomas of the larynx, colon, and ovary). Chapter 2 explains why specificity is an elusive criterion and depends importantly on exposure levels and host characteristics.

Temporal relationship (cause must precede effect) is an essential criterion. It is problematic only insofar as it is often difficult to sort out the time frame, particularly when there is a long latency until the effect is observed.

The biological gradient of exposure and response is very useful and compelling when it appears. The absence of an exposure-response relationship does not necessarily rule out an association, however. Chapter 2 explores this issue in greater depth. In many toxicological systems, response changes with exposure level, and greater levels of toxicity may obscure the expression of more subtle effects, usually by increased lethality. In epidemiological studies, this is less often a plausible explanation for failure to observe an exposure-response relationship.

Biological plausibility is often problematic as a criterion. The carcinogenicity of asbestos remains unexplained biologically, for example. Many strong associations prove elusive in the laboratory, such as the association between arsenic exposure and lung cancer. As well, science remains in doubt about many of the host defenses and adaptive mechanisms that alter outcome.

Coherence of evidence is a strong criterion, but it assumes a relatively thorough knowledge of the problem. Hence, this criterion (and often those for consistency and biological plausibility) is almost always useless in the early stages of investigation of a hazard, regardless of how compelling the evidence for an association initially.

Experimental or collateral validation may involve toxicological demonstration of a similar or comparable effect, but may also extend to experimental epidemiology (involving a controlled intervention among human populations usually in the form of clinical trials) or, less convincingly, quasi-experimental studies (evaluations following interventions without strict control). In effect, this criterion is an attempt to sidestep what is usually seen as the inability of epidemiology to directly test cause and effect.

Reasoning by analogy is one of the weaker criteria. It is actually closely akin to the criterion for coherence of evidence and contributes little by itself. Analogy is very useful in generating hypotheses and theoretical constructs but is invalid as a means of proof. Certainly the analogy between causes and mechanisms of pneumoconioses (a type of occupational lung disease) would not have predicted the carcinogenicity of asbestos. For proof of causation, however, empirical evidence is required; analogy is only a step along the way.

Different Hill criteria have different effects on interpretation. The criteria for consistency, biological plausibility, coherence, and collateral validation tend to be very conservative. They require a degree of documentation that epidemiology cannot guarantee (because of bias and confounding factors in the case of consistency) or cannot provide alone (as in the case of biological plausibility or collateral validation). These criteria make the Hill guidelines a “scientific” rather than a “public interest” test (see Chapter 7). However, control of a new or newly recognized occupational, environmental, or consumer hazard often cannot wait until the scientific profile is complete.

In short, the Bradford Hill criteria are important as guidance but are absolutely not conclusive. Indeed, his 1965 essay (in the *Proceedings of the Royal Society of Medicine*) is often misread. Hill did not believe or assert that these criteria were definitive. In fact, he wrote: “I do not believe ... that we can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we can accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help

us to make up our minds on the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?" There is no formula or easy route to interpretation and the inference of a relationship of cause and effect.

## **COMMON METHODOLOGICAL ISSUES AND INTERPRETATION**

The fundamental problem in etiological investigation in epidemiology is to identify which statistical associations are probably causal and which are not. Epidemiology, by itself, only identifies statistical associations: counts that are consistently similar or consistently different, frequencies that mirror or are the inverse of one another, trends that go in the same or opposite directions. The two sets of data show some pattern of relationship that implies an underlying linkage. Most often, in occupational and environmental epidemiology the findings being compared involve a measure or category of exposure and a measure or category of response (see Chapter 1). The core problem of most epidemiological interpretation in occupational and environmental epidemiology is to decide whether an apparent association reflects chance alone, bias, a cause-and-effect relationship, or some other relationship involving a shared cause.

### **Ruling Out Chance**

The first major issue is to determine whether a finding is likely to be due to chance alone and therefore unlikely to represent a causal effect. (Note that this proposition is worded to avoid the implication that at this early stage of analysis it is even possible to consider that the finding represents a causal effect. That comes later.) Apparent associations can be due to chance alone, and this possibility is precisely the reason that inferential statistics is used to evaluate the results. Statistical analysis is based on the assumption that there is a possibility of random error whenever a measurement is taken, as of an exposure, or chance can come into play, such as taking a sample to represent a

population or set. Random error arises simply from imperfection of measurement and chance occurrence.

Random error occurs when there is no factor or force that consistently causes a measurement or frequency count to be over- or underestimated compared to the true value, and is therefore called “nonsystematic” error. Inferential statistics deals with this problem tolerably well using traditional standards of performance developed in the early twentieth century, mostly by the statistician R. A. Fisher. Fisher considered the possibility of making a certain type of mistake, that of concluding that a finding is true when it is really due to chance alone (Type I error), more of an obstacle to the advancement of science than the alternative mistake, which is concluding that a finding is not true when it really is (Type II error). He therefore contrived what he considered to be a tolerable probability of error on the basis of chance alone of one in twenty replications of the same experiment (the basis for the  $\alpha$  probability of  $p \leq 0.05$ ).

While the conventional  $p$  value has served experimental science well, it is highly problematic in epidemiology. The conventional  $p$  value may be overly stringent in some cases in which the hypothesis is intrinsically difficult to test, and not stringent enough in others, as when there are multiple comparisons to be made, although in the latter case statistical corrections are usually applied. Today few epidemiologists would seriously argue that a  $p$  value of 0.055 can be summarily dismissed as “not statistically significant,” especially when there is ancillary supporting evidence (such as an exposure-response relationship or a similar finding in another study), or that a  $p$  value of 0.045 is definitive evidence of significance when it is unsupported by other findings. As a consequence, epidemiologists have moved away from rigid use of “ $p \leq 0.05$ ” as the criterion for statistical significance and now prefer a calculated confidence interval, which also provides more information and places the point estimate (the calculated value that represents the finding) in context.

The second major issue is that a high level of random error reduces the efficiency of epidemiological studies and the chance of observing what is really there. Biostatisticians after Fisher concurred

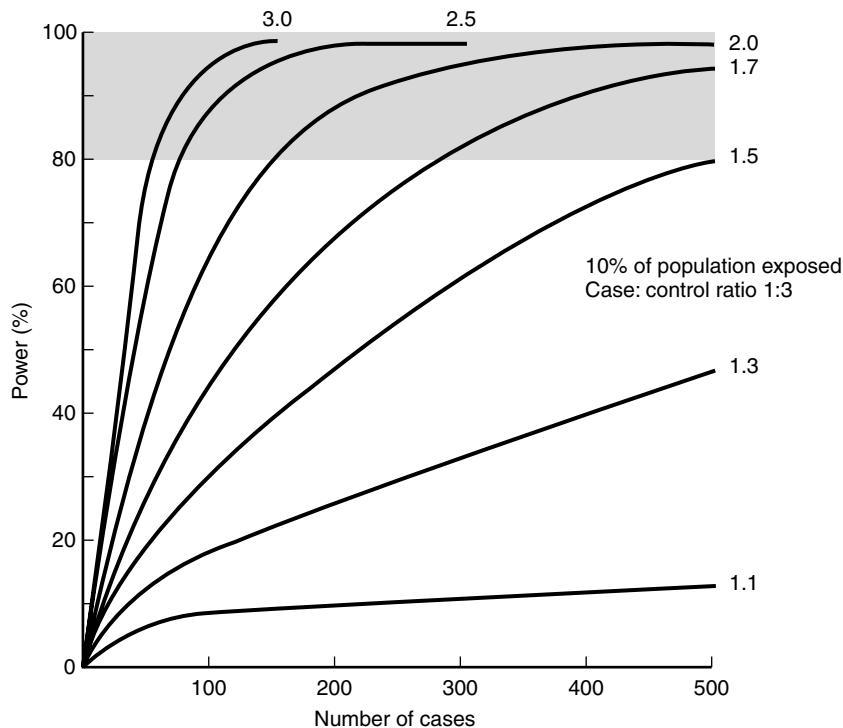
that a Type I error was more of an obstacle to the advancement of science than the alternative mistake, concluding that a finding is not true when it really is (Type II error), assuming, reasonably, that a “true fact” would always be rediscovered, but a “wrong fact” would be perpetuated and could mislead scientists for a long time. They therefore tolerated a much greater risk of error ( $\beta$  represents the probability of making a Type II error; values vary widely but 0.4 is not unusual). However, they also realized that the probability of making a Type II error had its own consequences, particularly the risk of missing a finding and rejecting a hypothesis prematurely.

### **Power and the Risk of Missing What Is There**

The risk of missing an elevated risk of a particular outcome that is really there is described by the familiar term “power.” Power is a statistical concept expressing the probability of detecting a true effect of a given size in a particular study. A power of 80 percent (or 0.80) means that if a disorder or outcome is present in the group in question at a given level higher than in the comparison group (say, twice as frequent), the study will have an 80 percent chance of detecting the difference as a statistically significant finding, or a 20 percent chance of missing it, although it is in fact present. A power of 0.80 is considered acceptable for most purposes, but the higher the power, the better. A power of 0.50 means that the chance of detecting a true difference is the same as flipping a coin and having it come up heads. It will be missed roughly half the time. A study with a power that low would not normally be a worthwhile undertaking if the outcome in question is all that was being examined, but in practice, prospective and cross-sectional studies usually examine many different outcomes, with different powers for detecting differences in each. Power then becomes a consideration in interpretation rather than study design, especially in evaluating the implications of a “negative” outcome, when no elevation over the expected risk is observed; if power is low (powers much less than 0.1 are not uncommon when data from studies designed for other outcomes are examined for rare diseases),

the absence of a finding in the study will not be convincing evidence that a true effect is not present, because it could easily have been missed.

Power depends on the number of expected cases, or background cases unrelated to the exposure of interest likely to appear in the population, and also on how great an effect the exposure is likely to have. Figure 3.2 illustrates the relationship between power and numbers and demonstrates two important constraints: (1) at some point, increasing the number in the study results in significantly diminished returns and may not be possible or worth doing; (2) it is very difficult to document a risk of magnitude less than 1.5 (50% elevated above



**Figure 3.2.** Power curves, relating power ( $1 - \beta$ ) to the number of expected cases in a study for a given effect size (expressed here in terms of the magnitude of relative risk); the hatched area represents a conventionally “acceptable” power of 80 percent, but most studies fall short of this.

background) without such large numbers that the study may not be feasible (unless based on a data set assembled for administrative purposes). This is reflected in the magnitude of the risk estimates that are expected. (Risk estimates are discussed later in this chapter.) Larger studies have greater power to detect an association, even with a relatively low effect size. Rarer disorders require that a much larger population be studied in order to yield the same power, usually even if the effect size is actually very high.

Formally, power is expressed as  $1 - \beta$ , where  $\beta$  represents the probability of making a Type II error. The major effects of random error are to reduce both the efficiency of epidemiological studies and the probability of finding an association where one truly exists. Power describes the statistical behavior of this random error. Evaluation of the power of a study to demonstrate what it is purporting to study is a critical step in the evaluation of so-called “negative” studies, which do not show an elevation in risk for a particular outcome.

For a given effect size, power is driven by number of subjects or years of observation in etiological epidemiological studies, and by the frequency of the outcome in absolute terms. For example, in a study of cancer risk, it will matter whether enough people are in the study to have a reasonable chance of seeing the elevated risk and whether a particular cancer is common or rare.

Power (actually,  $\beta$ ) is an important variable in equations to derive minimum or optimal study size. When a study is designed, a value of  $\beta$  (usually 0.2) is put into the equation to derive the number of subjects needed in the study to detect an effect of a particular size, measured by the risk estimate. The study design targets a main effect that reflects the purpose of the study, not every conceivable effect that could be examined. A study seeking to detect a risk that is doubled will require many fewer subjects than a study seeking to characterize risk at a relative risk of 1.5, which is generally small by epidemiological standards (except for environmental epidemiological studies of large populations using existing data sources). In practice, power is usually calculated after the fact for a study that has already been done, in order to evaluate a negative result (a finding of no elevated risk) for a different outcome, one that is usually much rarer

**Table 3.2.** Approximate Statistical Power to Detect Excess Risks Comparing Two Cohorts of 500 Workers Each, Followed for an Average of 20 Years

	Modeled On	Effect Size	Power
Disease 1a	Lung Cancer	2.0	0.84
Disease 1b	Lung Cancer	1.5	0.40
Disease 2	Brain Cancer	3.0	0.07
Disease 3	Myocardial Infarction	1.5	0.99+

Magnitude of the relative risk is the ratio of the cumulative incidence of a disease (or death) among the group of interest compared to a reference group, which in this case is another group of workers not sharing the exposure in question.

than the main effect. In etiological epidemiology, it is common for studies to look at multiple outcomes or for studies of a single outcome to be compared. Power may be quite different for the various outcomes and among various studies. For example, if a study was designed to look at heart disease, a very common health outcome, it is unlikely to have enough subjects to look at myeloma, a relatively rare cancer.

Table 3.2 compares typical study designs appropriate for the investigation of occupational diseases under three common sets of assumptions.

Disease 1 is modeled after lung cancer. Assume a relative risk (of one group over another) of 2.0 (double the risk in the “exposed” or high-risk group compared to the other), an incidence of the disease in the comparison group of 100/100,000/year (typical for a population of males in North America of mixed ages), and 100 percent case ascertainment (no cases missed or lost to discovery), an optimistic assumption. The power of such a study to detect the elevated rate would be a respectable 0.84. By comparison, if the true relative risk of Disease 1 were only 1.5, the power for that condition would only be 0.40. In other words, a difference in relative risk of 0.5 (equivalent to 50 percent increase in risk, or one quarter of the total risk of a doubling) means the difference between a probability of missing the association under 20 percent of the time and 60 percent of the time.

Disease 2 is modeled after brain cancer. Assume a relative risk of 3.0 (a high effect size but appropriate for some occupations), an incidence of 10/100,000/year (based again on North American figures), and 100 percent case ascertainment (not unreasonable, because the symptoms are often conspicuous and require surgery or specialty consultation). The power of such a study to detect the elevated rate is very poor, only 0.07. The study would be worthless in searching for association with this disorder, although if an association were indeed found, it would be noteworthy.

Disease 3 is modeled after myocardial infarction. Assume a relative risk of only 1.5 and an incidence of 300/100,000/year, because heart attacks are common and occupational exposures in industry are unlikely to raise the level of individual risk much higher. The power of such a study to detect the elevated rate would be quite high, over 0.99.

Thus, a study of two groups of 500 workers followed for 20 years would be expected to detect elevated rates of relatively common disorders even when the risks are not greatly elevated, but it would not be expected to detect such differences in rare disorders unless the risks were very high. Such calculations are routinely made before major studies are begun.

When the power of a single study is not sufficient to be confident of the result, a methodology called “meta-analysis” is often used. In meta-analysis, the risk estimates for several studies are combined, weighted to account for the relative power by sample size. The result is supposed to be a more reliable estimate for a larger, synthetic population that nominally consists of all the subjects in the individual studies. (“Pooling,” a related method, involves combining the data and then recalculating the risk estimate.) However, the outcome of meta-analysis in practice has been disappointing. For example, past efforts at meta-analysis in the literature on firefighters did not successfully identify several cancers for which later cohort studies provided strong evidence for an increased risk. Meta-analysis of several small studies on the effect of beta-blockers after myocardial infarction suggested considerable benefit; a randomized clinical trial to test this showed the opposite and had to be stopped. Reanalysis of existing published

studies, with all the limitations of that method, sometimes confuses the picture, rather than clarifying it. Because power, used in etiological research, describes the probability of detecting an elevated risk, it is predictable that studies of low power will most likely miss an association in the case of a rare disease. Including negative studies in a meta-analysis may merely be diluting the risk estimate, which may be more accurately reflected in studies that cluster around the true risk. Furthermore, the direction of most biases and power consideration is toward missing a true elevation, and combining studies does not overcome this limitation. Meta-analysis, in this view, is therefore barely useful when sufficient studies of reasonable quality do exist, and will certainly not be of any assistance when data are lacking or unreliable. Pooled studies with large populations but limited resolution have not fully resolved these issues, either. Meta-analysis is a helpful tool, but only that. It does not overcome gaps in the literature.

A related problem is “publication bias.” “Negative” studies may not be published because journals are generally reluctant to clutter their pages with studies that appear to be uninformative. (This is a strong argument for a public repository of negative studies, particularly for clinical trials.) This so-called publication bias results in an overrepresentation in the published literature of “positive” studies, which can bias meta-analyses performed from published studies. However, this is a factor only when the meta-analysis is examining an outcome that was related to the main findings of the paper and therefore affected its chances for publication. Meta-analyses, especially in occupational and environmental epidemiology, are often performed on studies in which the outcome of interest is secondary or incidental to the main hypothesis of a paper. In such cases, publication bias is not likely to give a falsely inflated risk estimate.

### Misclassification Error

A corollary to the problem of power is misclassification error, which occurs when a characteristic of the subject is important to the association under study (such as exposure history, age, or diagnosis)

but is gotten wrong or is wrongly recorded in a more or less random way. For example, a worker's job information might be inaccurate as taken off the death certificate, or the length of service on the job might be estimated from incomplete or secondary records and may be off by a few years either way. This type of error is called "misclassification." Misclassification error is relatively benign and varies depending on whether it is "nondifferential" or "nonsystematic" (more or less random and not always in the same direction). The effect of misclassification error is almost always to reduce the estimate of risk, making the association less visible and potentially obscuring it. The effect of nondifferential misclassification bias is simply to make the study less efficient, reducing its effective power.

Random error is generally benign in its effects, if at times inconvenient, and is relied upon by epidemiologists as the basis for statistical inference. Indeed, if it were not for the random distribution underlying uncertainty, informative statistics could not be applied.

### **Ruling Out Some Other Cause**

"Factor" is used loosely here to refer to an underlying tendency or force influencing the distribution of the findings and causing bias, as well as influencing other, nonapparent risk factors and outcomes, complicating the results.

Besides an effect of chance, which is ruled out first, an association may also appear in the data because the findings are biased. Bias is a systematic error in which the finding, usually a risk estimate, is consistently and systematically under- or overestimated because of some intrinsic problem with the study design or its implementation. Bias is a much more serious problem than random error, but it is intrinsic in epidemiological research; it is difficult to conceive of any study in the real world without some bias.

There are at least two dozen named sources of bias (the eminent epidemiologist David Sackett once compiled a catalog of them), including common errors that arise from selecting the study sample in a way that favors subjects with certain characteristics over others

(sampling bias), the tendency to remember things more completely after an illness or an event draws attention to their potential importance (recall bias), and misattributing characteristics of the subjects in a systematic way (differential misclassification error or misclassification bias). A critical part of interpreting epidemiological findings, especially with respect to causation, is to search for bias that could give spurious results.

An association may also occur because the findings are linked by a third factor, or still more factors, that correlate with both the risk factor under study and the outcome. This creates the illusion of a relationship or distorts the magnitude of the true relationship. For example, there may be a single, hidden cause that actually drives both the level of exposure and the outcome. The most common example of this in occupational and environmental epidemiology is socioeconomic status. People who are poorer and less educated tend to work in more hazardous jobs and are more likely to smoke and to smoke more heavily when they do, and tend to live in certain neighborhoods that are less affluent and more exposed to pollution. They are also more likely to develop a number of health conditions, such as heart disease, even when the amount of smoking is taken into account (“adjusted”). In this situation, a “social epidemiologist” would point to low socioeconomic status as the “root cause” of both the lifestyle difference and the health outcome. An etiologically oriented epidemiologist, on the other hand, would point to the material effects of lower socioeconomic status—greater likelihood of smoking and other behaviors associated with lifestyle in the social class, greater likelihood of employment in hazardous occupations, and residence in a more polluted environment—as causal, perhaps in a multifactoral manner (see below), with socioeconomic status being a proxy, or “surrogate marker,” for the root cause.

Some occupational epidemiologists, including Laura Punnett and others, have called attention to the sterility of these interpretations and developed the idea that occupation, socioeconomic status, and lifestyle are inextricably linked, or “embedded” in social role and history. In this view, study subjects have clustered, interdependent

characteristics reflecting the reality of their lives and work—both their individual choices and the social structures into which they fit. These characteristics represent different aspects of real life rather than discrete risk factors that happen to coincide. Root causes exist that drive intermediate risk factors, which move in tandem. This view is a modern update of the approach of “social epidemiology,” which in its application to occupational epidemiology and disease risk traditionally emphasizes social factors, employment, and class differences.

The more traditional view in epidemiology is to treat other risk factors linked to both cause and effect as “confounders” acting along parallel but different pathways. The phenomenon called “confounding” occurs when a third factor is associated with both the measurement of exposure and the outcome of interest. A classic example is the relationship between alcohol ingestion and lung cancer, which appears significant until smoking is factored in; in fact, people who drink heavily tend to smoke, and the association is actually mediated by smoking. In occupational epidemiology, smoking is often a confounder of greater or lesser significance. Although confounding is usually considered to be unfortunate in a study because it interferes with the interpretation of the data for the main point of interest, the confounding relationship may well be more important and interesting with respect to public health implications than the factor under study. Confounding factors are true risk factors that are related to primary causes, but they are not the risk factors of primary interest in a given study. One epidemiologist’s confounding factor is another’s main effect.

Confounding factors are not the same as background noise, nor are they the same as the numerous competing influences that affect a person’s opportunity to be exposed to a hazard or that may modify a person’s health outcome. These competing influences are likely to be distributed more-or-less randomly and are usually more-or-less adequately dealt with by conventional inferential statistics. The problem that usually proves most troublesome in epidemiological analysis is accounting for true confounding factors—those characteristics that are associated with both opportunity for exposure and

outcome. For example, acceptance of cigarette smoking has increasingly become a characteristic linked to social class in recent years. Smoking rates are much higher among blue-collar workers than white-collar workers. Smoking is also closely associated with risk of lung cancer. Therefore, any study of risk factors for lung cancer among foundry workers, who are in a blue-collar occupation, must account in some way for the contribution of smoking before concluding that any workplace exposure plays a role. Usually, this is done by selecting a relevant comparison group with matching characteristics. However, it can be demonstrated (as was first demonstrated by Tony Fletcher, using data on foundry workers) that in order to account for observed elevations in risk, the difference in smoking habits between the workers and the comparison group would have to be so high as to be implausible or even impossible (in his example, more than 100 percent of the foundry workers would have had to have smoked in order to account for the observed excess risk of lung cancer). The general lesson here is that smoking can be a confounder, but in real-world occupational epidemiology, confounding by smoking is usually overestimated.

It is a fundamental concept of epidemiology that “main effects,” the association of interest, will demonstrate consistency in association, whereas other complicating associations will be variable in strength. Two problems with this concept are: a relatively weak association may be lost among stronger complicating associations, and bias may mimic consistency in driving an association. Also, the very real risk factor being considered as a confounder may act in ways other than simply driving a competing pathway of effect in parallel with the main effect of interest.

Cigarette smoking may play other roles besides confounding. It may interact with workplace exposures (the best-documented examples include asbestos and radon), magnifying the cancer risk further. Competing mortality may obscure the association, because certain subjects who might have died eventually from smoking-related lung cancer may have died earlier from workplace exposure-related lung cancer, and vice versa. It is therefore just as great a mistake to

dismiss an association summarily, on the assumption that it is the result of confounding due to smoking, as it is to assume that smoking plays no role.

There are various ways to remove the influence or to assess the magnitude of confounding in practice. One is to adjust for the confounding factor, as one would for age or other characteristics. However, such methods depend on the availability of some measurement of the confounding exposure or risk factor. This is often unavailable because the confounder, by definition, was not the main interest in the study. The confounder may not have been considered in the study as carefully as the exposure of interest, may not have been incorporated into the study design, and may not even have been anticipated. One method that does not depend on the availability of data on the confounder is to assess the strength of the association between the exposure of interest and the outcome as the measurement of exposure becomes more accurate or precise by using subgroups in which the exposure of interest is more specific and subgroups in which the exposure of interest is greater but not much else is different. As the exposure of interest is isolated and increases in intensity, the main effect should strengthen and the effect of the confounder should weaken.

Occasionally, the exposure may be protective, preventing an effect that would have been caused by something else, and the frequency of the effect declines with increasing exposure. This is a causal relationship of a different type. For example, Parkinson's disease is known to be associated with exposure to pesticides (although it is not clear which pesticides and under what circumstances), and the available evidence suggests that the association is probably causal. However, there is also an inverse association between Parkinson's disease and cigarette smoking, which is causally protective against the disorder.

### **Getting Closer to Cause and Effect**

When chance, bias, linkage, and confounding have been ruled out by various means, a cause-and-effect relationship may then be seriously

considered. The exposure may be a true cause of the effect, such that the frequency of the effect rises with increasing exposure. The next step in assessing the likelihood of causation, explained in more detail below, involves a more detailed analysis of evidence internal to the study and from collateral sciences such as toxicology. Because of the complexity of the causal relationship, epidemiologists are usually very careful not to declare cause and effect prematurely.

A cause, in the sense used here, is a factor that contributes to the likelihood that an outcome will occur. In biomedical sciences such as toxicology, it may be defined as the event or condition that sets into motion a particular mechanism or chain of events that without interruption always leads to the outcome, but this is not how the term is used in epidemiology. In daily life as well, one may speak of “cause-and-effect” relationships as if there is one cause for every effect and as if an effect necessarily follows the presence of a cause. This is too rigid to be useful in etiological epidemiology and especially in cancer studies, where the mechanisms are complicated and may be influenced by numerous external and internal factors.

A cause is identified in an epidemiological study by documentation of a risk factor, indicator, biomarker, or exposure (often through imperfect measurement) related to the cause that increases the probability that something will happen, but may not guarantee it. This is a stochastic, or probabilistic, definition, not a mechanistic definition. There is a certain probability, or odds, that a step will occur, but no certainty.

The causal chain may not be direct or simple. The exposure of interest may well be just one of many factors that contribute to the risk of the disease outcome, as in the familiar case of smoking and cardiovascular disease. Such complex webs of causation are called “multifactorial” (sometimes “multifactorial”). There may be many contributing causes and other factors modifying risk up or down, but usually only a few practically important “drivers,” such as smoking and serum cholesterol with respect to cardiovascular disease, against a backdrop of hereditary factors.

Causation is therefore a more complicated concept when applied to epidemiology than is normally encountered in clinical medicine. Human populations are vastly more complex than experimental systems and are subject to numerous influences on health, behavior, and social adjustment. The aggregation of people into populations subject to epidemiological study both allows these complexities to be sorted out statistically and obscures many characteristics of the individual. In the presence of numerous complications, clear associations demonstrated in epidemiological studies are remarkable observations, suggesting three explanations: an effect, a statistical chance event, or a false finding reflecting a bias in the study method. If the latter two “fake” outcomes are excluded, the association that remains is not necessarily causal in nature.

Causation refers, in this context, to the risk factors or exposures that initiate the process leading to the health outcome. The concept is similar to that of etiology in clinical medicine but without the implication that there must be only a single cause. The concept of causation in epidemiology requires that the risk factors bear a “causal” relationship in that they either establish necessary conditions or set into motion a mechanism that results in the outcome.

A useful distinction can be made between causes that are risk factors in an epidemiological sense—in that they increase the probability of an outcome that is not certain—and those few that are invisible precisely because they are intrinsic components of the mechanism that produces the effect, such as oncogenes or pathways of the metabolism of procarcinogens. The latter “component causes” are more profitably considered means to the end rather than initiating events. Exposure to these component causes cannot be controlled because they are intrinsic; they may, however, be modified in such a way as to slow or prevent the action of the mechanism from producing the outcome. In this sense, when deciding to take a drive, the act of getting into a car and turning the ignition key are “causes” of a trip in an automobile (in this case directly related to a primary cause, the reason for the trip). The engine, drive train, and auto body are parts of the mechanism, but they are not causes in the sense that they initiate a sequence of events.

It is also not useful in this particular context (compared to causation assessment in medicolegal terms) to speak of a cause as being either “necessary” or “sufficient” or both, because causes may be interchangeable in the mechanism, or they may interact. For example, exposure to either cigarette smoking or asbestos individually is known to result in lung cancer in a roughly predictable probability. Exposure to both vastly increases the risk beyond that of the summed probabilities of either alone, suggesting a substantial interaction (see Chapter 10). However, most workers who have been exposed to either or to both do not develop lung cancer, although they might if they lived long enough and were free of other risks to their lives. A few unlucky people who neither smoke nor are exposed to occupational carcinogens such as asbestos do develop lung cancer regardless of their low risk, although this is very uncommon. In this example, neither asbestos exposure nor cigarette smoking is necessary, sufficient, or predictable in an individual case as a cause of lung cancer, but the association is clear and these factors truly are “causes.”

This example also illustrates the fallacy of trying to apportion the contribution in individual cases of multiple causes. Is the interaction in the case of asbestos and cigarette smoking one of asbestos enhancing the effect of cigarette smoking, or vice versa? In several exercises, authoritative investigators have attempted to estimate the proportion of cancer “caused” by various classes of external influences and have almost invariably concluded that smoking and diet are major causes of cancer in the population, and that occupational and environmental exposures contribute much less. It may well be true that control of smoking is the single most effective approach to the control of cancer risk that is now available. However, the apportionment of the relative contribution of various causes of cancer to overall cancer incidence or in a single case assumes that their effects on the underlying mechanism are separable and individually discrete when they clearly are not. It is a useless exercise to apportion causation on the basis of risk estimates for factors in (relative) isolation because the mechanism is intrinsically always interactive (see Chapter 2, section on carcinogenesis). As Kenneth Rothman put it in his landmark book *Modern*

*Epidemiology*, “it is easy to show that 100 percent of any disease is environmentally caused, and 100 percent is inherited as well. Any other view is based on a naive understanding of causation.”

## **STUDY DESIGN IN OCCUPATIONAL EPIDEMIOLOGY**

This section discusses study design from the point of view of interpreting the significance of a study for OEM, not from the point of view of the investigator tackling a scientific problem. It emphasizes particular aspects of concern to consumers of epidemiological information, including OEM physicians.

### **Occupational and Environmental Epidemiology**

Occupational and environmental epidemiology are sister disciplines but have differences in practice. Many epidemiologists do both. Historically, occupational epidemiologists have tended to cross over to air pollution studies on the environmental side more often than environmental epidemiologists who worked on water and air pollution have crossed over to occupational studies.

#### ***Occupational Epidemiology***

Studies in occupational epidemiology are usually concerned with the attempt to identify a specific hazard in the workplace and to quantify the degree of risk it presents at various exposure levels. They are usually searches for causes but often are efforts to document the extent of problems, the effectiveness of interventions, or the factors that modify the outcome. Many studies in occupational epidemiology overlap in methodology and approach the approach of social epidemiology, particularly those related to stress and working conditions. There has been a trend in recent years to evaluate the effectiveness of various treatments and approaches in injury care, disability management, and prevention in the workplace. The methodology of

these studies is essentially the same as those of clinical epidemiology (with exceptions to be noted) or intervention studies in the general healthcare and prevention literature, with the usual addition of cost and productivity data.

Often, a hazard cannot be characterized directly, and evidence of an unpredicted or suggestively high incidence (excess) of an outcome is required to provide leads for further investigation. This usually takes the form of broader survey studies conducted at the level of industries, plants, or occupations. Occasionally, studies of occupation-related risk are undertaken in the general population, usually to determine the overall burden of occupational disease.

Exposure-based studies are designed on the basis of known exposures and seek to document the effects on workers selected for their exposure at various levels and the opportunity to quantify personal exposure accurately. While ideal, exposure-based studies are often impossible to achieve. Exposure data are usually missing for years past, and the exposures that are the best candidates for association with a given health outcome may not be known in advance. Unless the investigator has legal authority for right of entry (as does the National Institute of Occupational Safety and Health), gaining access to the right population of workers is difficult and often impossible. Because it is usually so difficult to assemble a group of workers with a known, consistent, documented exposure to a particular hazard, studies are generally based on industries, individual plants, and occupational groups. Exposure is then measured and documented as it is observed or inferred from the jobs the workers do.

Industry-wide studies have an advantage in that they are potentially able to examine all or most occupational categories in the industry. Such studies are difficult to organize and very costly. They require cooperation from employers and in most cases from unions or other representatives of workers. Cohorts of workers in the industry can be assembled from employer or union rosters for fairly complete coverage. Pooling data from many employers and plants, however, risks diluting evidence for an effect in some location or classifications where the risk is higher.

Studies based at specific plants have many of the same advantages but usually involve many fewer workers and therefore have low statistical power beyond the most common health outcomes. Employers are, of course, highly sensitive about permitting studies to be done in their plants by outsiders.

Studies on occupational categories depend on the accuracy of identification of subjects. Usually, such studies are done from death certificates, cancer registries, or other data sets in which occupation is entered as part of the data obtained. The accuracy of this information is often questionable in the case of death certificates, because there is a common tendency to "upgrade" low-social status occupations in reporting by next of kin and because some jurisdictions report the last occupation held before death, whereas others report the usual occupation held during life. Seldom is more than one occupation recorded, even for persons who may have had very extensive occupational histories and numerous exposures. Occupational titles may not reflect the duties actually performed by a worker on the job and may be misleading with respect to exposures that may be encountered in one industry but not another, though that other industry may use the same occupational title. (Chapter 15 discusses the problems inherent in taking the occupational history in greater detail.)

Occupational epidemiology rarely uses two types of study designs. Ecological study designs, in which group characteristics are associated with group-level outcomes, are of very little interest because they are not very informative. When job-specific data are available, there is no point to such designs because a much more targeted analysis can be done. Randomized clinical trials are very rare in occupational epidemiology and are usually inappropriate in the workplace. Randomization within a department or workplace or among injured workers or claimants for workers' compensation creates divisions among employees or claimant groups, introduces unequal treatment (which may be grounds for legal action), and removes all incentive to participate from the reference or comparison group, especially if no other intervention is offered. Thus, intervention studies in occupational epidemiology tend to examine outcomes

of employers, departments, or entire plants with different characteristics, or they collect occupational information on participants in clinical trials that are based on diagnosis rather than occupation.

Occupational studies are prone to certain types of bias that are not often found in environmental studies. The most important are the “healthy worker” effects and migration or retention bias.

The “healthy worker effect” (there are actually two but they may act together) is a bias seen in prospective (cohort) studies in which the experience of an entire working or exposed population is compared to the general population instead of to another working population. Bias is introduced into the comparison because the working population is generally healthy and has fewer individuals who are at high risk of death or disease because of personal risk factors compared to the general population, which contains many members who are chronically ill, institutionalized, disabled, or not working for some reason that may increase their risk of dying or getting a disease. Prospective (cohort) studies use the relative risk as the primary estimate of disease or mortality risk. The relative risk is a ratio between the observed deaths in the working or exposed population and the expected deaths derived from a comparison population. If the comparison population is the general population, the effect of the bias is to increase the denominator, therefore reducing the estimate of risk and making it seem that the workers or exposed subjects are healthier or less likely to die than would be the case if they were compared to a group of true peers. The magnitude of this effect is quite large in older studies, with typically 10 to 20 percent fewer deaths recorded for workers than would be expected from the general population. However, the magnitude of the healthy worker effect seems to be diminishing over the decades.

The healthy worker effect arises from at least two sources of bias. These tendencies may operate in opposite ways over time, but the net effect is to make it appear that workers or exposed subjects have less risk of death or disease than they do.

The first source of bias is a form of recruitment or selection bias, although it is inherent in the study population and out of the control

of the investigator. People are discouraged from seeking employment if their health is poor. Those who apply for jobs are often screened for fitness for duty (see Chapter 18), which includes a medical evaluation. As a result, those who are given employment are more likely, at the time that they are hired, to be healthy or to have only minor disabilities that do not predict future disease risk. As they age and continue on the job, they may experience new or aggravated health problems, and over time the difference narrows between their health status, in the aggregate, and the average health status of people in the general population their own age. For many conditions, however, fitness at the time of hire predicts outcome when the worker is older, and so this healthy worker effect persists into and beyond retirement for some disorders, such as cardiovascular disease. It is not as strong for cancer, adjusting for smoking.

The second source of bias is, again, a form of selection bias over which the investigator has no control, but it reflects patterns of retention. When people who are on the job experience a health problem, they are more likely than their co-workers to leave the job as their condition gets worse. Thus, over time the workers who remain on the job are on average healthier than the workers who leave, and that much healthier than the general population. This form of bias holds for any chronic disorder likely to interfere with work, such as diabetes, asthma, and arthritis. This form of bias has several effects on the outcome of epidemiological studies. Over time the worker or exposed population becomes, on the one hand, less and less like the general population as subjects who develop chronic health problems drop out, and, on the other hand, the worker or exposed population converges with the general population on risk for some conditions, such as cancer.

When workers develop a work-related illness, such as occupational asthma, they are likely to leave the workforce and will be lost to cross-sectional studies. (This is why the prevalence—the frequency of cases among current workers—of asthma tends to be low in dusty work and among firefighters, even though the incidence—the frequency of new cases—may be high.) Such workers will be missed

on cross-sectional studies and may be hard to find or trace in cohort studies. Out-migration may also complicate other studies because workers may take other jobs, forget to mention their prior work, and the connection between their previous work and disease risk may never be made. It is very difficult to trace former employees, so out-migration from the workplace is a formidable challenge to investigators whose methods are based on surveys.

Finally, there is a reverse trend sometimes seen in occupations where there is a strong tradition of mutual employment protection, as among firefighters. When a worker cannot keep up, due to age, reduced fitness, or minor disease, that worker may be reassigned to other duties but kept on the force; or other workers may informally do the hard work or keep the worker on the sidelines when things get difficult. Reassignment would normally be documented and therefore could be taken into account in the exposure assessment. Informal protection is concealed, often to protect the weaker worker's respectability, income, or seniority, and so could be a source of bias.

These tendencies are becoming less and less strong, for several reasons.

The health of the general population is improving, particularly with reductions in cardiovascular risk. Workers therefore start out with less of an advantage than in the past. People are also living longer, so there is more time for the health risks of older workers or retirees to converge with survivors of the same age in the general population. Mortality and disease risk at advanced ages show a "survivor effect," such that mortality rates and cancer risk rise less markedly than in the "young old."

On the other hand, the workplace has become much more diverse (in terms of gender, age, and ethnicity), and social policies such as the Americans with Disabilities Act (see Chapter 18) have made it possible for workers with health risks and disabilities to be accommodated in their jobs and to stay at work. Policies on access, work accommodation, improved workplace design (including the popularity of ergonomics), and the progress of effort-sparing technology have made it possible

for workers with disabilities to compete for employment on a near-equal footing. Work is now less of an indicator of fitness than in the past.

### ***Environmental Epidemiology***

In general, studies in environmental epidemiology are usually concerned with documentation of the harmful effects of a more widespread or dispersed exposure. They are almost never performed to assess the effect of purposefully designed interventions, although they sometimes take advantage of intervening events and evaluate changes after new standards have been introduced. One seminal study on particulate air pollution tracked mortality in the community before, during, and after the shutdown of the emissions source, a steel mill. Many studies are undertaken to assess ill-defined or suspected environmental risks in a particular community.

Four types of studies are much more common in environmental epidemiology than in occupational epidemiology. All have much greater appeal to members of the public than to experienced epidemiologists and are often undertaken in response to public pressure rather than for scientific or professional public health reasons.

“Cluster investigations” are a search for a cause of a disease, usually cancer, that is perceived as in excess in that community. They are so named because they are usually mounted in response to awareness in the community of what is thought to be an unusually large number of cases of a particular diagnosis clustered in time or space. The usual study design appropriate for this problem is a “case-referent” study, described later in this chapter. Cluster investigations in occupational epidemiology have occasionally identified a new hazard (the most famous being vinyl chloride and the rare cancer hepatic angiosarcoma), but workers are a population with a high preexisting, or *a priori*, probability of exposure, making causal relationships for this population much more likely as an explanation for an observed association than in the community. Clusters in the community almost never uncover new hazards and rarely identify any underlying cause for the apparent

cluster itself, as documented by Ray Neutra, who conducted or oversaw hundreds of cluster investigations in California. Most reported clusters are not what they seem to be, and on closer inspection, they show unrecognized heterogeneity in the diagnosis (usually the type of cancer), patterns inconsistent with exposure or latency, and inaccurate health reporting. Those few that represent true clusters, meaning that the cases occur much more often than expected in time and space, are still not necessarily evidence of the effects of a common cause. By their nature, random events are not uniform; if they were they would not be random. A scatterplot of random events naturally looks clustered to the eye, and where those clusters occur it is tempting to look for a reason to explain the apparent but deceiving differences in frequency that look like aggregation. Clusters come to attention because they are unusual, not impossible. They may well occur by chance alone, and usually do, as the inevitable extreme event that just happened to occur in that location. However, though there may be an apparent cluster in a community, there are also many, many more clusters of other diseases that could have occurred in the community and did not, and many clusters of the same disease that could have occurred anywhere else in the world but did not—all by chance alone. Clusters are common, difficult to explain to the public, and usually unrewarding to pursue. Most large state and local departments of public health have protocols for screening reported clusters—beginning with a determination that the diagnoses are accurate and homogeneous—before committing any resources to their investigation.

Community environmental health survey or health assessment studies are undertaken to determine whether a community has a disease distribution or pattern that is consistent with an environmental risk or exposure. Calls for community environmental health assessments are usually driven by residents' concerns over a visible hazard (such as a nearby power plant), a known exposure (such as groundwater contamination by a gasoline plume, which is very common), a pending or proposed legal action, or the vague feeling in the community that something is not right. Like clusters, these

less well-defined situations can be screened for plausibility. For example, a study of a community near a landfill or hazardous waste site usually makes no sense if there is no intact pathway of exposure of residents to contaminated groundwater or other demonstrable means for the pollution to reach the residents. The most appropriate means of conducting such a study would normally be a case-referent study, but the much weaker cross-sectional study design (described later) is almost always done instead, often in response to community pressure. Community residents and activists often try to document their suspicions by conducting surveys or amateur epidemiological studies of their own, resulting in data that are usually uninterpretable and contacts that introduce bias into any subsequent studies. Sometimes they find clusters, which, as noted, are quite common if one searches for them. Community health surveys are often requested by community representatives or activists not only as a means of pursuing the cause of a perceived problem, but as measures for environmental justice, so that disparate health risks can be corrected. Officials and politicians often concede to these demands because it is the path of least resistance and has superficial plausibility. These studies are almost invariably placed under the effective control of a community steering body, which may nominally be organized as an advisory committee but in practice exercises political power and a veto on anything deemed unpopular or intrusive. Conducting community health assessments in this way raises all sorts of issues downstream that cause further controversy. Because the engaged community perceives the health assessment as a response to its own unique problem, as the way to document its situation in support of a future legal or political action, and as a means to validate what it thinks it knows already, it tends to be highly resistant to the idea that other, similar communities need to be studied as comparison populations; intolerant of negative results; manipulative in matters of study design (often with all good intentions); and impatient with systematic data collection. The study itself becomes a community issue and the object of struggles for control.

Ecological association, the weakest form of epidemiological study design, is much more common in environmental epidemiology than in occupational epidemiology, although it can also be found commonly in preliminary studies of lifestyle risk factors (usually to generate hypotheses), epidemiological studies conducted by “medical geographers” (investigators in academic geography who specialize in spatial associations) and economists (who sometimes extend the methods derived from econometrics into the study of disease outcomes). Because it is a basic study design, ecological association is described in detail later, in the subsection to follow on basic study designs.

Time series are longitudinal sets of data that track a dependent or outcome variable over time, for example, mortality in a given population—such as a city—over days, weeks, or months. Time series analyses can be seen as a prospective ecological study design in which the unit of analysis is the population, reflecting how a community responds over time. Time series has been used with particular success in air quality studies, for example to examine the relationship between ozone or fine-particulate air pollution and death from cardiovascular or other causes. Combined with other data, such as a time series on air pollution and another on weather, time series analysis can be a powerful tool for making correlations between health outcomes and changing environmental conditions. Most time series are heavily adjusted statistically in order to account for or smooth out seasonal variations and to disentangle the effects of closely related environmental factors, such as summer season, temperature, and ozone levels. It is usual for time series from several years to be averaged and smoothed in order to derive the strongest possible “signal” against the background “noise” of random or irrelevant variation. The approach works best when the environmental influence under study, such as air pollution, is widely and evenly distributed, such that exposure is roughly homogeneous. This condition is met well with air pollution, not so well with water pollution, and not well at all with occupational exposures, which is why it is rarely used in occupational epidemiology. When the condition is not met, time series are subject to the same serious drawbacks as ecological study designs in general.

## Descriptive Studies

Descriptive studies present the essential historical data of what happened, to whom, when, and where. The data are usually presented without a reference group or comparison other than benchmarking from the literature. When there is a reference group, it is not usually closely matched, because data have not been systematically collected with analysis in mind. Descriptive studies are often case studies of programs, the experiences of institutions, or observed trends in morbidity or mortality. They are not designed to test inferences.

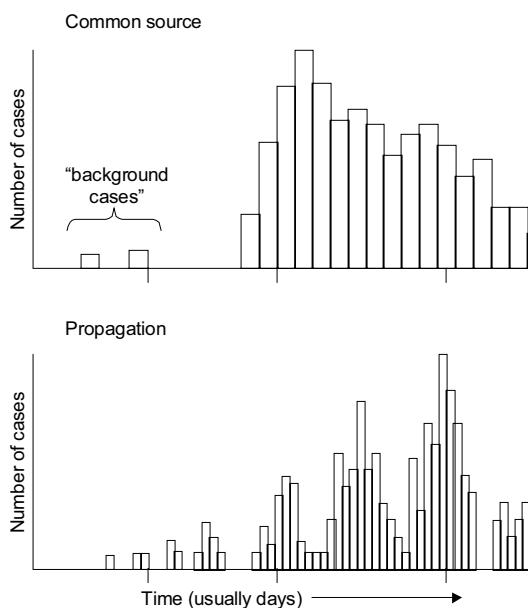
Descriptive studies have their place. They are useful in describing an event (such as a disease outbreak), program evaluation, healthcare management, and hypothesis generation. They may be essential to planning research, for example, to ensure that there will be a sufficient number of cases to support a more sophisticated study.

An example of a descriptive study would be the description of a registry that lists who developed cancer, the age of the patients, their respective occupations, and their respective locations of residence. The statistical summary may be essential to a cancer registry evaluating its services and may be very helpful to a hospital that is considering the best investment in patient care and staffing. The value of the data would be limited to that institution, however, because the data would reflect “numerator” trends (for example, how many cases did the cancer center see), as opposed to denominator trends (how many cases were there in the community as a whole and what was that institution’s share of the patient load). The descriptive data would say nothing about whether referral patterns were changing or disease rates in the community were changing.

By definition, descriptive studies are not conducted to test a specific hypothesis, so they are very limited with respect to causation analysis. However, they can be used in certain situations to make inferences when certain patterns appear.

When there is a disease outbreak, plotting the number of cases against time yields the “epidemic curve.” The shape of the epidemic curve can be used to infer the nature of the outbreak: either

“common source” or “propagated,” sometimes “transmitted.” When there are few or no initial cases and a sudden increase over a short time period, followed usually by a slower and more or less steady decline, the epidemic curve is consistent with a “common source”—a single location, vehicle, or exposure that affects many people at once. This is what one expects to see with most environmental exposures. The pattern characteristic of propagation is when there are a few early cases followed by an upsweep to a peak, which then either falls off (usually rather abruptly) or is followed by subsequent, usually larger peaks and valleys, called “waves.” This is the characteristic pattern of communicable, or “transmissible” diseases, where there is transmission from one person to another. Figure 3.3 compares these two patterns.



**Figure 3.3.** Epidemic curves, showing differences between the “common source” pattern (abrupt increase in cases, often against a background of a few sporadic background cases) and the “propagation” pattern (sweeping increase to a peak, then rapid decline often followed by successive waves).

Knowledge of the timing of an exposure, such as an incident of water pollution or the introduction of a communicable disease into the community, allows calculation of other useful information, such as incubation period, population susceptibility, and, for infectious agents, the all-important variable usually expressed as  $R$ , the average number of new cases that result from transmission of the agent from a single case when the disease is spreading.

### **Basic Inferential Study Designs**

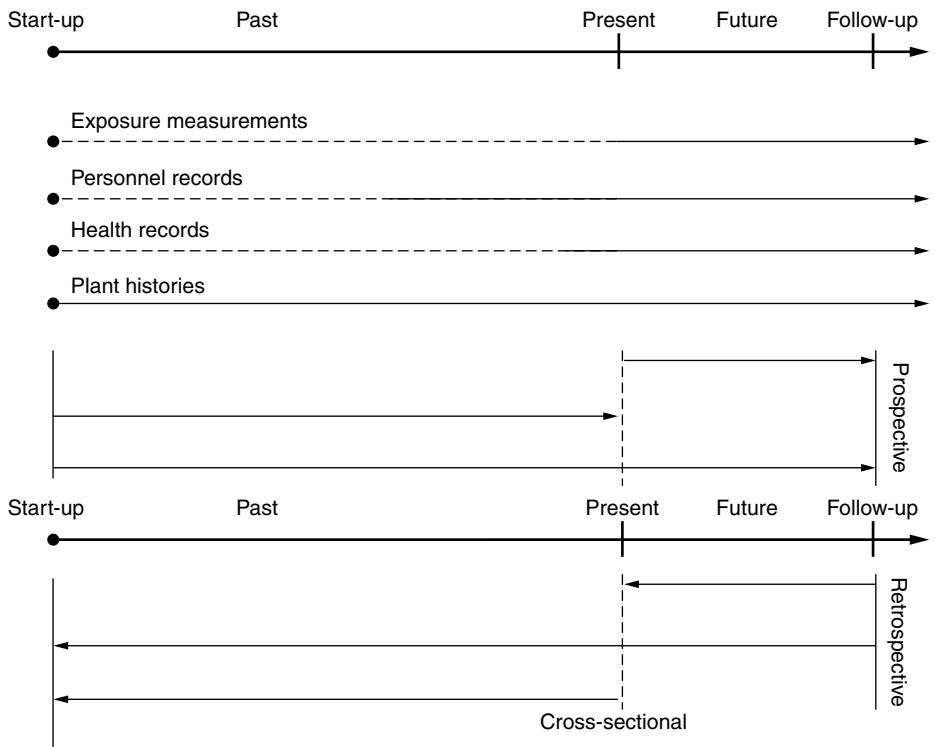
The ideal design for a proposed study makes the most efficient use of the available data; has a high level of statistical power; and, within the constraints of the available data and access to subjects, has the most adaptability to investigate different outcomes in the most informative ways.

There are four basic inferential study designs in occupational and environmental epidemiology, illustrated in Figure 3.4:

- Prospective, also known as longitudinal or “cohort” design
- Case-referent, also known as case-control or retrospective design
- Cross-sectional, also known as “prevalence” design
- Ecological association design

#### ***Prospective Studies***

Prospective, or cohort, studies follow a group of people (a “cohort”) over a period of time and record their outcome experience as a group against a comparison or reference group. Prospective studies can start in the present and follow the cohort forward into the future, or they may start in the past, if the cohort of workers of interest can be identified by records such as personnel records. The latter is called a “historical” prospective or cohort study and may end in the past after some period of elapsed time, in the present at the time of the study, or in the future after the study has begun.



**Figure 3.4.** Timeline for types of epidemiological study designs and the records used to support such studies.

Prospective studies allow for the investigation of many health outcomes at the same time and yield a reliable risk estimate based on the group's total experience. A prospective study design is best suited for etiological investigations in which the exposure of concern or putative cause can be identified and quantified at the beginning of each subject's exposure, and the subject can be tracked at least until the outcome appears. This period of time embraces both the onset of action of the first cause and the appearance of effect, documenting both with greater accuracy than other methods. When a prospective study is undertaken on a new cohort, exposure can be measured for the purposes of the study and so done more accurately, avoiding data gaps and analysis problems later.

In historical prospective studies, the cohort of workers or subjects of interest and a cohort of control, reference, or comparison subjects without the exposure or with minimal exposure are identified by records such as personnel records. Historical prospective studies usually rely either on exposure information obtained for other purposes, such as regulatory compliance (which may be incomplete) or inferred exposure, using values or classifications thought to be typical of the job (see Chapter 4). Groups, or subgroups within a larger group, may also be compared for differences related to exposures, work practices, or other characteristics.

The reference group or reference population provides a benchmark against which to compare the risk of the cohort. Much of the interpretation of cohort studies rests on determining whether the reference group or population is appropriate and to what extent the comparison is biased. Conceptually, a reference or control group is a second cohort with enough difference in exposure or risk characteristics to be a useful comparison, but sharing enough characteristics with the study cohort to be relevant. Ideally, and most efficiently, a “control” group (implying a quasi-experimental design) would be identical in the frequency and distribution of the characteristics of the cohort of interest, except that it would lack the single exposure or risk factor of interest. That is usually unattainable in practice, although in occupational epidemiology the reference group can often be matched fairly closely. In environmental epidemiology, health outcome rates from one community with a documented or suspected exposure are usually compared to rates from another, unaffected community that is close enough and similar enough to the study community to be relevant. Often, reference groups are drawn from unexposed or the least exposed subgroups within the cohort. In occupational epidemiology, this group is usually office workers, which inevitably raises issues of socioeconomic status and other confounding factors. In environmental epidemiology, this group is usually people living in another community or nearby who are not exposed to the hazard. Inevitable differences in age and other relevant factors are adjusted in the analysis.

If it is not possible to assemble a comparable group of control or reference subjects, a reference population may be used instead. Data from the population from which the cohort is drawn can be used as a benchmark to compare the outcomes to national, state, local, or industry-wide rates. As has been noted, in the discussion of the healthy worker effect, these comparisons are less satisfactory in occupational epidemiology than in environmental epidemiology because the general population consists of many people with health or other problems who cannot work, and lifestyle-related risk factors are associated with job and social status.

Finally, when there is sufficient variation, or “gradient,” of degree of exposures and frequency of outcomes in a large population, prospective studies can be analyzed by regression techniques that effectively use less-exposed and unexposed individuals among the subjects as internal controls for each risk factor, making very efficient use of the data.

Prospective studies yield incidence rates, which are statements of the frequency of an outcome for a population size over a period of time (for example, a cancer risk of 45 cases per 100,000 per year). The risk estimate is expressed as either a frequency or a ratio of frequencies (usually a standardized mortality ratio, explained below). Prospective studies have the great advantage of allowing the calculation of incidence: the frequency of new cases appearing over time. This is a fundamental and very useful measurement in epidemiology; it allows calculation of the cases attributable to the exposure or risk factor of interest.

The risk estimate for a prospective study is reported in comparison to the reference group or population. It is expressed as a ratio of the number of cases or incidence rates (the ratios are the same)—those observed in the numerator and those expected in the denominator—the expected number being what would be predicted for the study cohort on the basis of the experience of the reference group or population. This is the relative risk, which is the rate or number of cases observed divided by the rate or number that would have been expected or predicted for the cohort based on the experience of the

reference group or population. For example, if the risk of a rather rare cancer in a particular group of workers was 12/100,000/year and it was 8/100,000/year in the general population used as a reference group, the relative risk would be 1.5. In occupational epidemiology and especially in the historical literature, the relative risk is often expressed as a standardized mortality ratio (SMR). The SMR is the ratio of the number of cases observed to the number expected for a reference or comparison population of the same age distribution (and usually sex), normally expressed on a scale in which 100 represents unity (or as a percentage without the percent sign). In other words, if the same group of workers had among them 18 cases of cancer, and applying age-specific rates to a population of the same size would yield an expected number of 12 cancers of the same type, then the SMR would be reported as 150. A relative risk of 1.5 and an SMR of 150 are the same, although the use of the two estimates may imply some differences in the way the data are adjusted. The number of cases attributable to the exposure, or which are presumed to have arisen out of the workplace, would be the difference between the number in the exposed and the reference populations, which would be 6; this is called the “attributable risk,” which, despite its name, is expressed as a number rather than a rate. It is usually more useful to know what fraction of cases are related to the exposure. Because these 6 cases represent one-third of the total cases in the work group, the attributable risk fraction, expressed as a decimal, for the worker group is 0.33. Attributable risk and the attributable risk fraction are used in environmental epidemiology to set public health priorities, and in occupational epidemiology to establish the odds of causation in various occupational groups.

The period of follow-up is critically important and depends on what is being studied. Most contemporary cohort studies follow their cohorts for as long as possible, preferably “to extinction” (a term of art meaning until the end of the subjects’ lives or as long as they can be traced), in order to capture outcomes, such as cancer, that may arise after retirement and at advanced ages. Retirement is a poor cutoff point for studies in occupational epidemiology because health

outcomes with long latency periods, especially cancer, often occur after retirement and so are missed. At the other extreme, studies of reproductive effects and pregnancy outcomes or of acute disorders such as infectious disease obviously do not require long periods of observation.

Cohort studies are very efficient in the use of data because the statistical analysis can make use of every person-year of experience. They are best when performed with groups for which personnel records are reasonably complete but can be performed on incomplete records if exclusion of the subjects did not occur because of some health-related characteristic, such as early retirement or exposure to a hazard. They allow division of the workforce into subgroups by job, location, or exposure level at one time (for example, early in the latency period expected for cancer) or over the duration of employment, if the records permit. Groups, or subgroups within a larger group, may also be compared for differences related to exposures, work practices, or other characteristics. However, prospective studies tend to be very expensive, and fewer appear in the literature today than in the past.

A variation in the historical prospective study is the proportionate mortality (or morbidity) study. A proportionate mortality or morbidity study is done when the investigator has no way to identify the original members of a cohort, usually because there is no access to employment records. The data available are death certificates, cancer registration, or other end points at the end of life or at the time of diagnosis. The proportionate mortality analysis compares the proportions of death or other outcomes by various causes out of the total. If, say, cancer at a particular site is more frequent as a proportion than expected, that means that either there were more cancers of that type or there were fewer of other types than expected. Proportionate analyses are done only when there is no alternative because of the ambiguities and uncertainties involved in interpretation. They tend to be more useful for rare conditions, such as cancers, constituting a small proportion that is insensitive to differences in the denominator, than for major causes of death.

### ***Case-Referent Studies***

Case-referent (also case-control) studies are sometimes called retrospective studies, but this term invites confusion with historical prospective studies and so is discouraged. They start with the disease or outcome of interest and look backward to determine the exposure to risk factors experienced by cases compared to a sample from the same population that did not experience the same outcome. Case-referent studies are less straightforward to conceptualize than prospective studies, so they are more difficult to explain to the public.

It is useful to think of case-referent studies as being similar to a historical prospective study conducted on an imaginary cohort, the subjects of which cannot be identified but who have in common potential exposure to a risk factor in the past. Only the cases from the imaginary cohort are identified in a case-referent study; the reference group is selected by a process of sampling. The members of the imaginary exposed cohort from which the cases arose cannot be distinguished at the outset from the unexposed reference, or control, population, but they are identified by gathering information on exposure. There are usually many more unexposed than exposed individuals in the population base on which the study is performed (which could be the community, patients in a hospital, or workers in an industry, for example). Because only part of the imaginary cohort is known, the study design is based on sampling the population base and seeing how many of the subjects in the sample share the exposures or other risk factors with the cases (and so belong to the imaginary cohort) and how many do not (and so belong to the imaginary comparison or control population). The reference group, therefore, is a synthetic population consisting of a sample of the population base with the same matched characteristics of the cases, so that age and sex, usually, are accounted for. Unless exposure was extremely common in the population as a whole, the reference group will be mostly a sample of the imaginary comparison or control population and will behave as an unexposed reference group. The objective is to compare how common exposure was among the cases with how common it was among the reference group.

In practice, cases of the disease or disorder in question are identified through medical records, death certificates, or other means. The referent cases, or controls, are usually drawn from personnel records, neighborhood residents, or hospital admissions, as appropriate to the study. One or (usually) more referent cases without the disease or condition in question are selected at random (or as close to it as possible), usually matched to a particular case of the same sex and close to the same age. Other characteristics may be matched, but too much matching, called “overmatching,” is technically undesirable because it prevents the investigator from assessing the contribution of the variables being matched and reduces the chance of finding important interactions and unexpected risk factors. The characteristics of interest (such as jobs, location, exposure, prior illness, cigarette smoking) are then statistically compared between the cases and the controls.

Case-referent studies do not provide direct information on incidence or prevalence, nor do they provide data that can be used to assess attribution. Case-referent studies use a measure of risk called the “odds ratio,” which is mathematically related to the relative risk but calculated in a very different way. The result is a statement of the odds that a case, rather than a control, will have a certain characteristic or history of exposure to a risk factor. For example, “even odds” means a ratio of 1:1, which means that a subject who developed the disease has the same chance of having a risk factor as a referent who did not develop the disease, which is unremarkable. An odds ratio of 2:1 means that the subject had twice the probability of having that risk factor, which is an association.

Case-referent studies are easier and usually much cheaper to perform than prospective studies, but they can only examine one disease or outcome at a time. (Of course, several case-referent studies can be conducted simultaneously on the same population.) They are possible even when personnel records are relatively incomplete, as long as the records do not vary in completeness because of relevant factors. For example, if the files of workers who left work for medical reasons or who left work on permanent disability following workers’ compensation claims were put in a separate file by the human resources

office, the personnel records might be biased against finding relevant medical conditions and would probably bias the odds ratio for many risk factors.

Although case-referent studies are inefficient in the sense that they do not use all the available information, they are robust in the sense that they are not usually much affected by minor errors or chance. They are easy to incorporate into larger prospective studies and are often used to examine interesting findings or to suggest associations in greater detail. Case-control studies are generally the most practical way to study relatively rare diseases, but they can also be used for common disorders; case-referent studies were the basic design for the early studies linking lung cancer to cigarette smoking. They can also be performed on subjects within a cohort, using cases within the cohort and selecting controls from among cohort members who did not get the disease of interest. These are called “nested” case-control studies and are often used to investigate associations that may not have been thought of when the cohort study was designed.

### *Cross-Sectional Studies*

Cross-sectional (prevalence) studies examine the frequency of disorders and characteristics in question among subjects at one point in time, usually comparing two or more groups. Cross-sectional studies provide an excellent picture of the health status of a group at that time and can provide useful information that can be used in the design of more definitive prospective or retrospective studies. However, they cannot reliably assess trends and have very limited value in evaluating etiology, except to generate new hypotheses.

Cross-sectional studies, like prospective studies, can examine many different outcomes at the same time. Cross-sectional studies compare two or more groups when done for etiological epidemiology, but they are often performed as descriptive studies on subject groups without a reference or comparison group when the purpose of the study is to characterize a current state of the population (for example, an opinion survey), to establish a baseline for surveillance

(see Chapter 5), or for planning purposes, such as assessing the need for health promotion programs.

Cross-sectional studies can be used to study a population in considerable depth, to establish the frequency of many disorders at the time of the study (prevalence), and to estimate rough incidence rates of easily recognizable disorders by asking about personal medical histories over the preceding few years. But because the most reliable information is that obtained for the present, cross-sectional studies suffer from many disadvantages when used alone. They do not accurately reflect the experience of the group over time. Such studies are relatively inefficient because they cannot identify subjects who will develop disorders in the near future and are not reliable, in most cases, in detecting subjects who have had less easily recognizable disorders in the past but are now well. They are particularly limited in providing information on disability-related occupational disorders because workers who develop conditions that interfere with their work are likely to leave work and will therefore be absent from the group at the time of the study. For this reason, prevalence estimates in cross-sectional studies of occupational groups are likely to be underestimates of disease occurrence in the workforce.

By themselves, cross-sectional studies are of limited usefulness, but in combination with other techniques they are immensely valuable. Cross-sectional studies are excellent first steps in the development of more comprehensive studies. When cross-sectional studies are repeated or subjects are followed forward into the future, the efficiency and utility of the study increases greatly as it approaches a truly prospective study. In an industry with low turnover, cross-sectional surveys at the beginning and end of a long period could provide valuable follow-up, as if those workers were part of a prospective study based on personnel records. Sometimes, as in studies of psychosocial stress or depression, cross-sectional studies in the form of surveys are required because the outcomes examined are not easily ascertained from other records, such as death certificates or medical records. Cross-sectional studies can also identify cases and generate lists of suitable controls for later retrospective studies.

Survey research, using questionnaires and other direct data-gathering techniques, can be thought of as a common form of cross-sectional study. When cross-sectional studies, particularly surveys, are repeated, they are called a “series,” and when they collect data over a period on the same population, they begin to approach the structure of a prospective study or a time series.

Cross-sectional studies, like prospective studies, can examine many different outcomes at the same time, but because they cannot embrace cause and effect they are intrinsically weak when applied to etiological investigation. Cross-sectional studies provide “prevalence” data—a measure of how many cases there are or the frequency of risk factors present in the population at that point in time. The prevalence found at the time of the study is sometimes called “point prevalence.” “Period prevalence” is a measure of how many cases are present (whether pre-existing or newly occurring) in the population over a relatively short period, such as a year. The risk estimate used in cross-sectional studies is the relative risk (observed risk divided by expected risk, which is derived from the reference or comparison group), but increasingly, regression methods yielding odds ratios have become popular. Survey research is a basic method in social sciences and uses the same analytical methods as cross-sectional studies.

### ***Ecological Association***

Ecological association studies take the approach of relating group exposure and consumption characteristics to group outcomes. They are conducted on secondary data sources, most often linking exposure data in a geographical area to health statistics. A common example has been to relate cancer rates in various counties to the concentration of industrial activity in each county, or disease rates in a set of countries to dietary consumption. These studies are useful in generating hypotheses for further study, but they are weak evidence of a causal association.

The principal reason for their weakness is a phenomenon called the “ecological fallacy,” which states that such studies cannot relate

individual exposure to individual effect among subjects, and therefore cannot prove that the subjects who sustained the exposure were the same ones demonstrating the effect. There may well be different communities within the geographical area in question with quite different exposure and risk profiles, and some subgroups in the area of study may differ from the majority in important ways. However, such studies cannot go beyond data for the overall area in level of analysis. Few communities are homogeneous, and one group or social stratum may be consuming the product or working in the industry while another is experiencing the health problem. Because of their reliance on data sources compiled for other purposes, ecological association studies using international comparisons are especially prone to misinterpretation and errors, because of differences in reporting disease, access to medical care, and availability of data on exposure.

Ecological association studies are relatively easy and inexpensive to perform because they can usually be constructed from data that has been collected for other purposes. The findings of such studies often suggest promising leads for focused study using more refined techniques, but they cannot prove an association.

Because they are limited to one point in time, ecological association studies also share many of the drawbacks of cross-sectional studies. The problem is compounded because the time frame for collection of data used in ecological association studies may not fit a causation model; for example, health outcomes in the present, such as cancer incidence, cannot be explained by current data on income, food consumption patterns, environmental exposure, or other health-related characteristics, because cause must precede effect and cancer has a long latency period. Current data in that case is only useful insofar as it might suggest the prevalence of risk factors in the past, and extrapolating back twenty years may be too speculative.

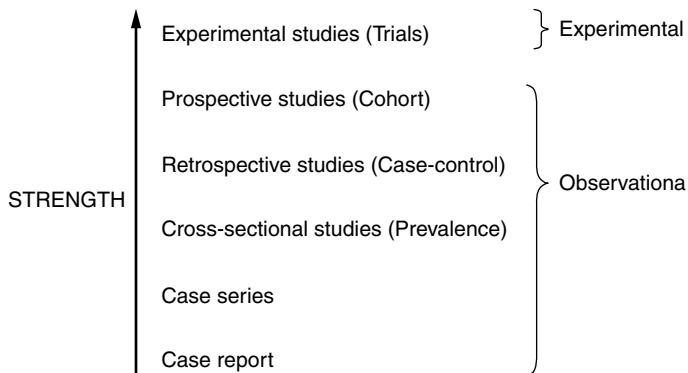
Unfortunately, because ecological association studies use essentially the same method as the regression-based and econometric approaches, which are the norm in economic research dealing with population-level effects and economic drivers, a number of economists

have entered the field of epidemiological research through ecological association studies without understanding their limitations for investigating etiology. For example, studies by economists have attempted to correlate the incidence of autism with indices of time spent watching television on a state or marketing survey level, without taking into account the intrinsic limitations, confounders, and differences in group-level, as opposed to individual, analysis. The result is noise and confusion without insight. Because ecological association studies are cheap, easy to do from available data, and easy to understand, they will continue to have popularity out of proportion to their validity.

### *Other Aspects of Study Design*

Study designs can be combined, such as by conducting a nested case-referent study within a prospective study, or by conducting a cross-sectional study followed by a long-term monitoring program to convert it into a prospective study. In recent years, conventional methods of analysis—for example, comparing an exposed to an unexposed group—have also been replaced by regression techniques, which make use of the full range of data for each variable and can more easily be used to study interactions among variables.

Figure 3.5 presents a conventional hierarchy of epidemiological study designs applicable to OEM, ranked in order of current perceptions of their intrinsic validity. The perception of the usefulness of the various basic study designs has changed with increased understanding over time. Case-control (or “case-referent”) studies were once thought to be a poor substitute for cohort studies, but they are now considered nearly equivalent in validity. The shifts in opinion have come about because of methodological research by a number of theoretically inclined epidemiologists, particularly Olli Miettinen of Finland (now at McGill University) and Kenneth Rothman (formerly of the University of Massachusetts). Also, the emergence of “clinical epidemiology” as the application of epidemiological methods to clinical research has provided insights into the value as well as the limitations of well-documented case reports and case series.



**Figure 3.5.** Relative strength of study types for proving a hypothesis.

There are other study designs that are important in OEM that are not epidemiological, in the sense that they do not lend themselves to inferential testing and do not conform to epidemiological methodology. These include case series and individual case reports. These investigations are most often used when the condition being studied is rare or unpredictable, usually in uncommon toxicity syndromes.

Case reports are often useful as educational illustrations of unusual situations or physiological mechanisms and may be all that are available in the case of rare disorders. Individual cases cannot be used as proof of associations; however, they may suggest an association. Case series are reports of several cases that share common features and an analysis of their similarities and differences; case reports are individual case descriptions. Case reports are not formal study designs in epidemiology *per se*, but they are important and necessary in OEM. Epidemiological investigations have to start somewhere, and for rare diseases, individual case reports and case series can contribute by suggesting the diagnostic criteria for a case and by raising hypotheses that can be tested by more formal studies. They may be all that are available for information on rare disorders, especially in clinical toxicology. They are often useful as educational illustrations of unusual situations or physiological mechanisms. Case series have the

advantage of larger numbers and at least provide a sense of the distribution of characteristics for that particular group of patients. Case series are subject to a strong referral bias because the characteristics of the cases being reported may not reflect those of all patients with the disorder—only those who happened to have reached the office or institution of the author, or who met the criteria of the investigators. Case series and reports may raise hypotheses that can be tested by more formal studies.

Quasi-experimental evaluation study designs are a set of methods for evaluating the outcome of interventions. They were developed largely for purposes of assessing educational innovations but have become widely used in evaluating healthcare innovations, health promotion interventions, and workplace interventions, more often of a psychosocial nature. The most familiar example is the “pre- and posttest” design, in which performance at a group level (such as productivity, absence, injury rate, expenditures, participation rates, medication compliance) is compared before and after an intervention or training program. Obviously, there could be other factors influencing performance, especially over a long period of time, so the “pre- and posttest” design is often supplemented with a control or comparison group—sometimes several with different characteristics—that benchmarks what would have happened had the intervention not occurred. From these basic building blocks, elaborate evaluation strategies can be constructed to assess interventions on many levels, from the individual to the organization. Collectively, these methods and their application are called “evaluation research” and are a foundation of social sciences research and, in disguise, business financial analysis.

Evaluation research can use any quantitative measure or qualitative method. Its statistical tools are common to epidemiological methods. However, the validity of the evaluation depends critically on how the evaluation is designed, what measures are used, what data has been collected during the course of the program, how accurate the data are, and whether all relevant dimensions of the evaluation are undertaken.

Usually, program evaluation is done for administrative or financial purposes, in order to show that a program is cost-effective or well implemented. This leads to practical problems in most business organizations, because in order to be instituted, a program has to have champions who value the program and are committed to it to the point that they would be embarrassed by failure. Organizational politics and human rivalry being what it is, these champions may also have attracted detractors who want to avoid spending money on the program, or who may have competing priorities and are looking for reasons to terminate it. For the group supervisor or manager responsible for production, the intervention may or may not be welcome in the first place, but evaluation almost always is not welcome, because it takes some amount of time away from production, it is almost always perceived as burdensome, and the results, even if fair, may not be favorable.

Evaluation research is best done by outsiders rather than sponsors or those engaged in running the program. This is because evaluation is usually done for a specific purpose and has consequences. Those most engaged in a program may seek to defend it and gloss over its limitations. Those who supervise it or pay for it may have an interest in seeing it terminated or cut back. There may be competing agendas to sideline the program, or it may be a favored project that nobody in the organization wants to see fail. Because of these possible agendas, it is best to bring in someone disinterested, who does not have a stake in the outcome.

## OUTCOMES AND RISK ESTIMATES

The selection of an appropriate outcome to study is a choice made largely on the basis of practical considerations: What is described in the literature? What is of concern in the community? What is discussed and mentioned in folklore? What can be measured? and What is most important for prevention? When numerous health outcomes are sought in any large population, a few will be found to be in excess on the basis of chance alone, and the proper interpretation of the findings may therefore be inconclusive.

## Outcomes of Interest

The outcomes of interest in OEM can be divided into four general classes: mortality (causes of death); morbidity (illness or disability); health-related behaviors (such as frequency of cigarette smoking); and subclinical findings (such as changes in pulmonary function) that are not in the category of overt disease but that indicate possible trends, possible mechanisms of action, and patterns of toxicological significance. Increasingly, subclinical findings, evolving diseases, and even levels of physiological adaptation are identified by biomarkers of effect (discussed later in this section).

Mortality is usually determined through examining death certificates, which are relatively easy for investigators to obtain because they are public records. Death certificates vary in their accuracy by location, circumstance of death, and diagnosis. They are more accurate for traumatic injuries and common cancers—because cancer requires a tissue diagnosis—and much less accurate for sudden death and mortality of the elderly, especially in nursing homes, for whom the cause of death is unlikely to be investigated. An important reason for this variation in quality has been the lack of rigor in specifying the cause of death by the physician who signs the death certificate, who in many cases is not the patient's treating physician. A historical reason has been the low rate of autopsy confirmation of cause of death and the practice of forwarding death certificates to offices of vital statistics before the autopsy results become available. Death certificates usually ask for an immediate cause of death, an underlying cause of death, and associated conditions. The underlying cause of death is almost always the entry of interest.

Morbidity is much more difficult to determine because both personal and occupational health records are less accessible, considered confidential, and frequently incomplete. Morbidity can be determined by direct surveys or measurements of the population, but this is difficult and expensive and, in the case of surveys, may depend on the subjects' faulty recall of events occurring before the time of the survey. Some diseases, such as tuberculosis, are of sufficient public

health interest to require mandatory reporting by physicians, although for most diseases underreporting occurs anyway. Special registries exist in most jurisdictions for cancer and birth defects. Workers' compensation data are difficult to work with and are often of questionable reliability because of their tendency to reflect administrative criteria, changing policies on the acceptability of claims, and inconsistency in acceptance of individual claims; but they provide the only available information for many types of occupational injury and illness.

Health-related behaviors on the part of workers often play as great or greater a role in determining the outcome than the occupational exposure in question. They are particularly important to characterize in order to correct for confounding factors in a final epidemiological analysis. Health-related behaviors (for example, smoking or alcohol consumption) are seldom recorded as completely or accurately as would be desirable in medical records, and are usually determined through direct questioning or from the medical record. These histories are prone to misrepresentation, always in the direction of social peer pressure, such as underreporting tobacco or alcohol use. Direct validation of the subject's responses is always a practical problem; especially because medical records are prone to similar reporting biases. For subjects who are deceased, recall by next of kin or surviving colleagues is sometimes undertaken but is often impractical.

Clinical testing can be determined by a survey, a direct examination, or a review of the medical records. There are obvious drawbacks to each: self-reported health information on surveys may not be honestly reported, or the subject may exhibit recall bias. Direct examination is subject to problems of precision and interlaboratory testing differences, which require standardized testing and performance of tests in a reference laboratory with elaborate quality assurance. A small systematic error may be of no great significance in clinical testing, but in a study of many subjects, perhaps thousands, it can lead to serious bias and invalidate the results. Medical records are rarely available for all subjects in a study and require such extensive confidentiality and other protections that they are not usually used. Medical records are

also subject to bias, as noted before, and the laboratory results, while adequate for clinical purposes, are not standardized.

Biomarkers are measurements, usually taken on body fluids, that characterize an individual subject biologically and toxicologically. Biomarkers are based on toxicokinetics, biological mechanisms, and genomics. There are several types of biomarkers, listed below:

- Biomarkers of exposure, which assess the individual subject's actual exposure from all sources. This may be a biological exposure indicator, such as blood lead, or an indirect measure such as DNA base-carcinogen adducts. This class of biomarker reflects the history of intake into the body as recorded by body burden or biological changes.
- Biomarkers of effect, which assess the effect on the body of the exposure and the level of toxicity, which is evolving. This may be a positive result on a screening test or a more subtle effect, such as depressed cholinesterase levels as an indicator of exposure to organophosphate pesticides. This class of biomarker reflects the history of the body's efforts to adapt or respond to the exposure.
- Biomarkers of susceptibility, which assess states or conditions that make the subject's body more susceptible to the effects of an exposure than would be the case for others or the same person at other times. This may be an inherited condition, such as rapid acetylation in drug metabolism or G-6-PD deficiency, or it might be an acquired condition, such as early changes on the spirogram predictive of accelerated loss of pulmonary function.

Biomarkers as used in epidemiological studies are generally collected according to a rigid protocol with stringent quality assurance. It is unusual for biomonitoring data to be abstracted from the clinical record, because of the more casual nature of data gathering, interlaboratory variability, and the difference in accuracy required for diagnostic tests compared to screening and monitoring tests. Without

detracting from the need for accuracy in clinical medicine, a measurement that may be just a little off, within a reasonable range, is usually still sufficient for purposes of diagnosis when there is a high *a priori* likelihood of disease. However, a level of error that would not interfere with diagnosis can severely bias the estimate obtained in an epidemiological study, especially one in which extrapolations are made.

### Estimates of Risk

Estimates of risk are measures used to compare the experience of the study group against a benchmark, usually that of the general population or a reference group. An estimate of risk requires an expression of frequency. For example, the risk of common cancers is usually expressed as the number of cases per 100,000 persons per year. The current rate of lung cancer (latest available rolling average, 2002 to 2005) is 70 per 100,000 per year among men and 41 per 100,000 per year among women.

The frequency alone has limited usefulness; it is mainly used in insurance, actuarial calculations, and planning for the provision of healthcare. It is generally more useful to compare the frequency to a relevant benchmark. The simplest way to express relative rates is as “odds,” which are basically the ratio of the frequency of the event occurring to the notional “frequency of it not occurring,” indicated by the ratio of people with the disease to the people who do not get it. The lifetime probability of developing lung cancer is 1 in 13 for men and 1 in 16 for women, meaning that the odds of not getting cancer are 13 to 1 and 16 to 1, respectively. This form of risk expression is generally more useful for the individual, but it tells nothing about whether the odds are high or low or how they compare with a different community or population.

More useful still is a risk estimate that compares the experience of subjects, usually the people exposed to the hazard of interest, to some relevant standard of comparison. For this, a suitable comparison population or benchmark is needed, which can be called the “reference.” Classical epidemiology does not make a distinction between comparison

populations and benchmarks. Even so, the distinction is of great practical importance.

The ideal reference population is a “comparison population”—a close match to the study population. A comparison population should be as close to the study population as possible, demographically and in the distribution of other risks, so that the effect of the hazard can be isolated to the greatest extent possible. It is similar to a control group in an experiment. One important feature of the comparison group is that it should have been studied in the same way, using the same methodology, in order to avoid bias. Statistically, populations are adjusted in the analysis for factors such as age, in order to make them comparable. The uncertainty in the risk estimate arises out of uncertainty in the frequency of the outcome or characteristic among cases, if the cases are considered as a sample of the total universe of people like them, and also from the frequency of the outcome or characteristic among reference subjects, if the cases are considered as a sample of the total universe of people who, like them, are unlikely to have the disease or would be unlikely to have encountered the hazard being studied, if the hazard is associated with the outcome. However, this is outweighed by the much greater similarity of the groups being compared, which makes comparability more reliable and revealing.

A benchmark is not the same thing as the experience of a comparison group. A benchmark is a relevant comparison that is known or understood not to be particularly similar to the study population and that is not treated in the same way as a control group would be in experimental science. Rather, a benchmark is a reliable measurement against which the risk of the study group can be compared in terms of magnitude or trend. The purpose of the benchmark is to take into account the general trends to which the study population is responding apart from the specific risk factor. For example, if one is studying the influence of cell phones on injury rates among bicycle couriers, there is no particularly good comparison population available consisting of young, diverse urban professionals who bicycle wherever they go but do not use cell phones. Attempts to isolate the distracting effects of cell phones would be limited to observing effects over time, correlated

with cell phone use. However, it would be relevant to know whether injury rates in general changed over that time, because couriers might just be reflecting general trends. One approach would be to compare injury trends among couriers with trends among the population as a whole, or within a relevant demographic, in order to rule out other explanations or influences that may have occurred at the same time. The most common benchmark is the experience of the general population, which is understood to be quite different from the study population, but which provides a stable reference point from which to compare rates and trends. Benchmarks are usually derived from much larger populations or from data gathered for other purposes, such as vital statistics or administrative data. When such a benchmark is used, the uncertainty mainly arises from the estimates coming from the study population, because the benchmark population, being much larger, yields much more statistically stable estimates. However, the use of a benchmark introduces issues of comparability, and the selection of the benchmark has to be thought through and defended. It is easy to rig a study by choosing an inappropriate benchmark. For example, in monitoring pulmonary function in workers exposed to a respiratory hazard, a set of normal values could be chosen from among the several that are available that are lower than other sets of data, making the numbers look better.

### ***Relative Risk and Related Measures***

“Relative risk” measures the frequency or rate of a disease, an event (such as death), or an outcome (such as disability) in a group of individuals who are exposed to a risk factor, such as a hazard in the workplace or environment, against the frequency observed in some reference group. In cohort studies, the preferred reference group is a similarly studied population with similar characteristics (demographically similar and perhaps living nearby or working for the same employer) who are not exposed to the hazard. This can be hard to achieve, for practical reasons. Often, epidemiologists use the general population of the area, state, or country as a benchmark.

In cross-sectional studies and surveys, if it is an occupational health study, the reference group is usually another group of workers; if it is an environmental study, the reference group comprises residents of a nearby community.

Rather than using crude or adjusted rates as such, relative risk is usually calculated by using the reference population—whether it is a comparison population or a benchmark—to generate estimates of how many cases would arise if the reference population had the same age, sex, or other relevant distribution factors as the study population: a procedure called “adjustment.” (Adjustment can also go the other way: the study population can be adjusted against a standard population, such as a national census.) This yields the number of cases that would have occurred if the study population were just like the reference population demographically.

Conceptually, this calculated number is the same as the number of cases one would expect to arise from the study population if the hazard were not having an effect and there were no other factors influencing its experience with respect to disease risk. For this reason, it is called the “expected” number of cases and is used as an estimate of the number of cases there should be. This number is then compared to the actual number of cases observed.

The calculation for a relative risk is simple conceptually but requires a great deal of data to achieve. It reduces to a ratio of observed to expected cases:

$$\begin{aligned} RR &= \text{rate in the study population/rate in the reference} \\ &\quad \text{population} \\ &= \text{number of cases observed/number of cases expected} \end{aligned}$$

The relative risk is usually expressed as a ratio and may approach zero (no risk) or infinity (if only subjects in the group get the disease). A relative risk of 1.00 means that the risk for the study population is the same as for the reference population, and there is therefore no difference. It is given as a point estimate and a confidence interval that covers the estimate with 95 percent confidence. It should be

noted, of course, that if the confidence interval does not contain 1.00, the risk estimate is significant at the level of  $p < 0.05$ . Relative risks are usually expressed in the following manner:

The authors of one important study of firefighters (Baris 2001) observed statistically significant excesses for colon cancer (1.51; 95 percent CI = 1.18–1.93). Nonsignificant excesses were reported for cancers of the buccal cavity and pharynx (1.36; 95 percent CI = 0.97, 2.14); for non-Hodgkin's lymphoma (1.41; 95 percent CI = 0.91, 2.19); for multiple myeloma (1.68; 95 percent CI = 0.90–3.11); and for lung cancer (1.13; 95 percent CI = 0.97–1.32). With >20 years of firefighting, the following cancer sites showed elevated risks: colon cancer (1.68; 95 percent CI = 1.17–2.40); kidney cancer (2.20; 95% CI = 1.18–4.08); non-Hodgkins lymphoma (1.72; 95 percent CI = 0.90–3.31); multiple myeloma (2.31; 95 percent CI = 1.04–5.16); and benign neoplasms (2.54; 95 percent CI = 1.06–6.11).

However, there are other ways to present relative risk. The most important in OEM is the “standardized mortality ratio” (SMR), which is used for death rates. The SMR was the most important index in occupational epidemiology historically, although it is now giving way to the simple relative risk. It is calculated in practice as follows:

The number of deaths from a specific cause and all occupations in the general population for each age group  $i$  ( $d_i$ ), over the population of individuals in the data set in the same age group ( $n_i$ ), yields the mortality rate. The product of this and the population of individuals in the given occupation  $g$ , falling into this age group ( $n_{g,i}$ ), is the age-specific expected mortality ( $E_{g,i}$ ). The sum of the age-specific expected mortality of all age groups, usually between 20 and 64 years of age if retirement is at 65, is the expected mortality for the occupation ( $E_g$ ).

The morbidity or mortality ratio is then given by:

$$\text{SMR} = \frac{\text{observed number of deaths/cases (O)} \times 100}{\text{expected number of deaths/cases (E)}}$$

The SMR, as it was commonly used, was basically an adjusted relative risk cut off at age 65 and expressed as a percentage in order to make it easier to communicate the implications of the study. A similar measure, the standardized morbidity ratio, was applied to illness rates. Today, percentages are not used as often in epidemiology, and it is recognized that cutting off the mortality at retirement age misses too many deaths that could arise from occupational risk factors.

If an investigator knows only how many deaths occurred for particular causes in a group and has no way of knowing what the mortality rate was (deaths per number of people per unit of time), the “proportionate mortality ratio” (PMR) is the appropriate expression of risk. PMRs are ratios of the fraction (proportion) of deaths (mortality) from a given cause in the group of interest compared to the fraction seen in a comparison group, usually the general population, with adjustment for sex and age distribution. PMR studies are well suited to the study of rare diseases, such as nasal cancer. They are poorly suited to the study of common causes of death, however. PMRs can be very useful in providing clues but are not definitive, because they are overly sensitive to two factors that cannot be known for certain from the available data: the death rates from causes of death other than the ones of interest, and the total mortality (death rates). If the overall death rate is excessively high or low, a PMR cannot demonstrate this to the investigator, because it is always 100 by definition. If the rate of a common disease is unusually high or low, it may “squeeze” or “expand” the PMR for another cause, because the PMR is based on proportions.

In general, PMR studies are more subject to distortion and error than cohort studies. However, the PMR closely approximates the SMR in situations where follow-up is complete and all deaths in a

**Table 3.3.** Comparison of Risk Estimates for an Occupational Cohort\*

Cause of Death	Observed Deaths	SMR	95% CI <sup>a</sup> of SMR	PMR	95% CI <sup>a</sup> of PMR
All causes	370	96	87–107	100	(Exact, by definition)
Infectious, parasitic	2	33	4–120	44	12–165
Malignant neoplasms (all)	92	127	102–155	124	103–148
Stomach	6	81	30–176	82	37–181
Colon, rectum	14	161	88–271	156	94–261
Lung	24	142	91–211	132	90–194
Prostate	8	146	63–288	138	70–273
Bladder <sup>b</sup>	4	316	86–808	328	130–824
Circulatory	157	103	88–121	102	91–114
Respiratory	21	95	59–145	100	66–150
Digestive	9	47	21–89	50	26–92
Genitourinary	8	132	57–260	146	74–289
External causes (trauma and poisonings)	49	66	49–87	80	64–99
Endocrine	3	47	10–137	49	16–147
Blood	1	102	12–370	102	26–407
Nervous system	4	74	20–188	80	30–212
Mental disorders	19	455	274–711	480	321–718

\* Firefighters in two cities in the province of Alberta, Canada, from 1927 to 1987, constituting 64,000 person-years.

a. Confidence interval for 95 percent probability, corresponding to mean plus or minus two standard deviations.

b. In this case, the PMR overestimated the SMR and produced a confidence interval that could have resulted in the finding being declared significant at  $p < 0.05$ .

cohort have been accounted for. Table 3.3 compares SMR and PMR results for one such study (the firefighter study described below) and shows that they are very close, but that the PMR occasionally exaggerates the risk in either direction compared to the SMR.

The proportionate mortality ratio (PMR) is calculated in the same way as the SMR, but the expected number of cases is calculated on the proportion, not the actual number, of deaths from the cause that would have been expected, rather than from a proportion based on the documented rates for the group and a comparison population. Proportionate mortality studies use death certificates and are not cut off at age 65. It is rare for proportionate analysis to be applied to other outcomes, such as the proportion of total cancers.

A proportionate analysis can be compared to dividing a pie. It is like judging the fairness of distribution of dessert to children around the dinner table by assessing the percentage of an already partly eaten pie given to each. Without knowing the original size of the pie, or how big a piece was missing in the first place, one can only look for gross discrepancies in who got what, not whether the children got the share of the pie they wanted. PMRs are strongest when the major causes of death are distributed as expected, where cardiovascular deaths show a PMR of 100, and where there is at least some evidence that the overall mortality experience is not grossly different from the general population. An SMR study, on the other hand, is like comparing the actual sizes of pies, or pieces thereof, by measuring their area or weight and determining which is larger or smaller and by how much. An SMR is a comparison of the number of deaths actually observed from a given cause to the number expected if the population of interest were the same as the general population or a comparison group. Statistically, it is a much stronger measurement. An odds ratio is like comparing pieces and determining which pie the piece is more likely to have come from.

### **Odds Ratios**

As noted, odds are one way of expressing probabilities. “Odds ratios” compare the frequency (probability) of characteristics in one

group compared to another. Because case-referent studies are based on the frequency of characteristics observed among cases and compared to reference subjects, without knowing more about the population the cases and referent cases came from, the only available risk estimate is the odds ratio (OR). The odds ratio is an expression of the frequency with which a given risk factor is observed among the cases compared to the frequency observed among the referents. One may recall that the relative risk used in cohort studies is derived from a ratio between two rates: the rate of disease in the exposed and the rate of disease in the reference group. The odds ratio, on the other hand, is a ratio between two less well-defined frequencies that approximate the rates at which a risk factor is found in the subject cases and in the referents.

$$\text{OR} = \frac{\frac{\text{number with characteristic among cases}}{\text{number of cases}}}{\frac{\text{number with characteristic among referents}}{\text{number of referents}}}$$

This is enough to identify differences in the frequencies of characteristics, such as the exposures, lifestyle risk factors, biomarkers, and genes, which is very useful in etiological epidemiology and applied prevention.

Odds ratios are actually another estimate of the relative risk in disguise, arrived at by a very different method in case-control studies and subject to greater uncertainty, but valid if the study is carefully designed. Although this is not the place to present the formal derivation, the odds ratio and the relative risk are mathematically related in such a way that as a case control study becomes increasing large and inclusive of the members of a cohort, however unknown or notional, it approximates the relative risk for the same cohort. This is another reason why contemporary epidemiologists do not consider the case-referent study to be inferior to or less exact than a cohort study—just a different approach with different trade-offs.

## RECORDS AND EPIDEMIOLOGICAL RESEARCH

Records are critical to occupational epidemiology, but are less important in environmental epidemiology. The existence, quality, and completeness of records set limits on what can be studied and the reliability of the data.

Five common types of records are generally used in occupational studies (Table 3.4):

- Exposure records
- Medical records of employees
- Personnel records
- Plant engineering records
- Death certificates, which are obtained from government sources

Figure 3.4 also represents the degree of completeness that can be expected for the four types of records usually available from industry, on a time line in relation to plant start-up, the present, and the future. The death certificate, of course, is a public record and should always be available with enough effort.

### Exposure Records

Exposure records are usually kept by occupational (industrial) hygienists in order to track measurements of chemical or physical exposure levels for workers. They are most likely to be complete if an OSHA standard applies, because they are needed to demonstrate compliance. Records may be kept for several reasons other than compliance:

- Routine area monitoring conducted to ensure safe working conditions and compliance with regulatory standards
- Special comprehensive occupational hygiene surveys conducted for a special purpose (often after acquisition of a plant by a large corporation or as part of the corporate response to an incident)

- Health hazard evaluations conducted to investigate a particular problem
- Personal exposure monitoring to determine the level of exposure in a given job, usually for purposes of assessing the need for or effectiveness of protection

Sometimes, the investigator will be fortunate enough to find exposure information that was collected using a research protocol, that is available in electronic form, and that is unusually accurate and complete. This is so rare as to be a once-in-a-career event. Usually, historical exposure data are a hodgepodge of measurements collected at different times, using different methods, with varying quality assurance. Such data can only give a crude estimate of exposure in a workplace or job and are mostly useful to define the range; they should not be relied upon as an accurate exposure assessment. The only common exception is noise exposure, which tends to be measured more often and more accurately than chemical exposures. Current occupational (industrial) hygiene data are more likely to have been collected systematically and entered into a database than in the past.

The few comprehensive surveys that were conducted systematically in the distant past were usually done at large plants by occupational hygiene teams from the corporate office of a major company or by local health and safety officers on orders from a corporate office assuming responsibility for a plant. These data are usually spotty in coverage, are often not completely documented as to the methods used, and are often difficult to retrieve from existing files. The exceptions to this generalization have been noise exposure, which is generally well documented, and asbestos exposure, which has been a major issue in some plants for many years and the subject of thorough searches in order to remove the material.

Whatever the limitations of these data for epidemiological purposes, they may have been adequate for the operational purposes for which they were gathered. Their limitations as applied to epidemiology do

**Table 3.4.** Records Used in Occupational Epidemiology

Type of Records	Type of Data	Locations	Uses	Likely Problems	Additional Information
Exposure	Occupational hygiene, routine monitoring, spot checks and health hazard evaluations, alarms	Corporate	Estimate exposure levels	Insufficient in past. Environmental monitoring data may not be representative of individual employees' exposures.	Surveys, monitoring systems, extrapolations to past (see Chapter 4).
Medical	Employee health status, demographics, risk factors (e.g., smoking), sickness, absence, disability evaluation, workers' compensation claims	Corporate or local	Confounders (e.g., smoking), hypothesis generation, validating diagnoses	Clinical data is variable in quality; incomplete records; nonstandardized data collection often omits smoking	Improved record keeping, health insurance (usually inaccessible), workers' compensation claims
Personnel	Date of hire and leave; job assignments; identify members of cohort	Corporate, local plant, union	Length of service, exposure assessment, demographic profile, medical leave	Old records are often incomplete and may be scattered among sites; retirees may not be in current records.	Direct surveys of workers and retirees

Plant history	Engineering records, technical modifications, process changes, new chemical hazards introduced, plant maintenance	Local plant or corporate	Exposure assessment, identifying upset conditions, identifying incidents, identifying locations of highest exposure, identifying time of introduction of new hazards	Seldom consolidated; poorly accessible; require technical skill in reading; minor changes may not be recorded; some records may be discarded	Interview retirees and "old timers"
---------------	---	--------------------------	--	--	-------------------------------------

---

not necessarily reflect carelessness or an attitude of neglect. Certain types of industry that have not been subject to regulatory attention may have had little reason to engage in extensive occupational exposure monitoring.

Chapter 4 addresses exposure assessment in greater detail.

### ***Medical Records***

Medical records prepared by the occupational health service serving the employer have an important purpose in clinical care, but they have limitations for epidemiological research. By themselves, they are rarely sufficient for epidemiological purposes.

Medical records are, of course, subject to numerous legal protections with respect to confidentiality and expectations of privacy. Although HIPAA (the universal acronym for the act governing confidentiality of medical records in the United States) is not always clear about how it applies to records in occupational medicine (especially as they pertain to workers' compensation), it is prudent to treat all medical records as if they were covered under HIPAA or similar legislation. Research using medical records must be reviewed by an Institutional Review Board (an authorized committee for the protection of human subjects and for ethics) if there is any intention of sharing the results outside the occupational health service (for example, for publication in any reputable journal). Standard precautions must be taken, for example, to preserve anonymity of the subjects and to ensure that information that could identify individuals is not made available, even within the organization.

Medical records may include information on personal health status, personal risk factors (such as cigarette smoking), incidents of injury or illness on or off the job, and demographic information. These data are gathered primarily for operational purposes, as documentation required under occupational health and safety regulations, for evaluation of claims under workers' compensation or healthcare insurance programs, for control of sickness absence, for determination of fitness to return to work, and for employee well-being and

health promotion. Information that might be contained in a medical record includes:

- Pre-placement evaluations (formerly pre-employment screening; see Chapter 18)
- Periodic health evaluations (generally limited to workers in high-risk situations or as mandated; see Chapters 5 and 18)
- By regulation or as mandated for safety-sensitive positions (such as commercial drivers, under U.S. Department of Transportation requirements)
- Return-to-work evaluations (fitness for duty; see Chapter 18)
- Workers' compensation reports
- First-aid or injury reports or logs
- Special records (such as respirator testing, audiometric testing, and pulmonary function testing)
- Optional records (such as health hazard appraisals or medical evaluations for health promotion programs)
- Exit evaluations (for those leaving employment)
- Safety committee minutes (for investigation of incidents)

Records collated for medical evaluation vary greatly in their completeness and specificity. Acquisitions and mergers have left many companies with a patchwork of record systems for their employees, with numerous inconsistencies and large gaps in coverage.

In recent decades, medical records have become more detailed and standardized. Early medical records, especially before 1980, often gave little detail and were often handwritten, creating a challenge for conducting historical cohort studies. In many community practices, such records are still handwritten and lacking in detail. Many employers only required a physician's approval or the physician's completion of a checklist for the worker to return to duty. Descriptions of symptoms were often sketchy and incomplete. In general, injury-related information was usually recorded in more detail in workers'

compensation claims, first-aid logs, or dispensary records. Personal risk factors, such as smoking, were rarely and inconsistently recorded until recently.

The usefulness of medical records for research purposes is limited unless the system was designed for standardized data collection and precise definition. Medical records created for occupational health purposes typically reflect the standards of clinical practice, where obvious abnormalities are sought, where much of the information is self-reported or copied from other records (often perpetuating inaccuracies), and where medical management is described in an abbreviated way in terms meaningful only to other health professionals, and not easily abstracted by research personnel.

Differences in measurement, terminology, and routinely recorded information may make little difference in the handling of a particular case clinically, but in an epidemiologic study, such differences could introduce bias. For example, if all workers at a particular plant were to be examined by the same physician or tested by the same laboratory, but employees at another plant were examined by another physician and tested by a different laboratory, there would almost certainly be bias. A physician evaluating workers at the plant in question might be unusually thorough in taking a history and writing down minor findings, and the laboratory used by that plant's medical service might consistently report white cell counts 5 to 10 percent below those of other laboratories because of slight differences in technique. If one plant had a somewhat higher level of exposure than others, the differences might easily but falsely be attributed to the difference in exposure levels. At the very least, there would be a problem in risk communication in explaining the finding.

Medical records are often inaccurate for lifestyle factors such as smoking, drinking, and exercise. Whether they are workers engaged in routine screening or patients being treated, people tend to underestimate the amount that they smoke, underreport their alcohol intake, and exaggerate their commitment to exercise. They will also rarely volunteer information on such risk factors unless they are asked. Because of the sensitivity of such information, occupational

physicians often do not ask, and this information is not available in many records. Smoking was commonly omitted from the record in the past, in part because employees (and some employers) think that it is none of the employer's business. If only workers whose history explicitly mentioned smoking in the medical record, in the absence of a direct question, were assumed to be smokers, many smokers would be misclassified as "nonsmokers." A study of a smoking-related outcome such as lung cancer could then be seriously biased; the net effect of such a bad study would be to artificially reduce the estimate of risk associated with smoking and quite possibly to falsely attribute an excess rate in a particular plant to occupational exposures because the risk would falsely appear elevated among "nonsmokers."

Smoking data, specifically, is often recorded on the request form for workers who undergo pulmonary function tests as part of a periodic health evaluation, but compliance with such programs is not complete, and not all workers are part of such programs.

Medical records must be used judiciously for research—only when the information collected is specific, consistent, and reasonably complete. Clinical laboratory studies cannot be taken at face value as outcome measurements, especially at values close to the normal range, because they may vary slightly in precision by technician, by laboratory, by equipment, and over time. For the diagnosis of an illness, this seldom matters much, but for epidemiologic studies such differences may introduce serious bias. Most problematic of all are physicians' assessments. Even identically trained clinicians vary in their ability to hear a heart murmur or the degree to which they may be impressed by "soft" signs in a neurological examination. As long as the inaccuracy is not great, precision does not matter much for purposes of clinical diagnosis or periodic health surveillance (see Chapter 5). The problem arises when one group is compared to another, which is the essence of epidemiologic research. If the findings are consistently or, more often, inconsistently over- or undercounted in one group relative to another, the comparison may be far off base. This is not to say that medical records are

useless for epidemiological research, however. They are invaluable for six purposes:

- To ascertain the baseline health status of entrants into the industry in order to document or rule out selection factors in the recruitment of workers (such as the exclusion of workers with conditions such as asthma)
- To determine whether the present workforce reflects the out-migration of persons who develop health problems during their employment
- To identify a subgroup of workers with particular characteristics (such as smokers) in order to study them separately and to look for interactive effects
- To identify a group of workers who develop a certain condition during the period covered by the study
- To identify workers who die or leave employment during the period covered by the study
- To identify suspicious patterns of disease reporting or prevalence for more detailed investigation

In the United States, information recorded before the 1970s is almost always less complete and certainly less standardized than data collected since because of the influence of the Occupational Safety and Health Act. Smoking or alcohol intake was rarely recorded in the past and is only now recorded in some companies, particularly those with active health promotion programs. The reliability of recorded information on alcohol intake in particular is open to question.

Medical records are gradually being centralized throughout industry, particularly as acquisitions and mergers result in plants consolidating their employee health programs. Some records are kept only at local plant offices or under the care of a community-based occupational health services provider. Data recorded for personnel at headquarters or in the largest facilities, where occupational health services are centralized, tend to be more complete and more consistent

with company guidelines than data reported for workers in smaller and often remote plants. A common problem arising from this is that field workers tend to be less well-described in the records than headquarters workers, who are of less interest in etiological investigation because they mostly work in offices.

A further problem is accessibility. Files are organized for case management rather than easy retrieval of data. However, at this time (2009) the use of electronic medical records is probably more common in occupational health services than in general healthcare. The most popular programs allow for information retrieval and basic analysis.

### ***Personnel Records***

Personnel records are essential to a prospective occupational health study because they are necessary to construct a cohort. However, they are not always easy to use.

Personnel records are kept for operational purposes: to coordinate staffing, establish seniority, make payrolls, and meet the responsibilities of taxes and reporting requirements. They are never designed for purposes of research. Linked with other records and data files, however, they are the key to an occupational health study, because they usually provide, at a minimum, four of the five essential elements required: identification of the worker (needed to link with other records), age, sex, and work history (or at least duration). The fifth element, health outcome, must come from elsewhere, but correct identification from the personnel record is required to establish the linkage with health information. For those reasons, the quality of the personnel records is often the single most important factor in determining whether a study is possible.

Access to the records of retired employees is critical, especially in the investigation of disorders with long latency periods. Some companies purged certain types of dated personnel records every five years in the past, although that is rare today. Others store them on microfiche or other hardcopy, which is hard to search and abstract.

Many simply store the paper documents in closets, warehouses, or “inactive record” rooms.

For smaller operations and companies that grew by acquisition of smaller plants, personnel records tend to be spotty and often scattered by location. A major limitation on the scope of personnel records is that they seldom cover the employees of contractors or subcontractors, some of whom are on-site working alongside regular employees. Work in the current economy is increasingly contracted out to vendors, other companies or individuals, and in such situations no health records are kept by the company letting the contract.

A major problem in studies in occupational epidemiology that do not rely on death certificates is the handling of retirees, although this is important for diseases with long latency. Tracing former employees is not easy for some companies, particularly those that have grown by acquisition. Pensioners may be readily identified if they are still receiving checks, but reconstruction of the past work history may be very difficult, in some cases possible only by a questionnaire or by interview of surviving retirees after they are traced to a current address. Some will be unable to respond. Nonpensioned retirees or persons leaving employment prior to eligibility would have no follow-up in current files and would be even more difficult to trace.

Job descriptions and location assignments are often hard to interpret from older records. Seldom is it possible to retrieve records by specific plant, job assignment, or workplace location, rather than by individual identifier, except for current assignments. Results of job descriptions and locations are often unavailable for distantly retired workers or those who left before earning eligibility for a pension, but they are usually readily available for the most recent retirees and former employees. In cases of interrupted employment—when a worker leaves an employer and is later rehired by the same employer—there may be discontinuity in the record, with the earlier work history lost.

The trades are a particular problem because their work is organized by project and they go from job to job with various employers, often

with periods of inactivity. Duration of employment is usually reconstructed by means of stubs or union records.

### ***Plant Engineering Records***

Of the four record types, plant histories are usually the most complete and detailed, reflecting major engineering changes, plant modifications, and process changes from start-up. However, these records seldom exist in the form of concise historical reconstructions, although there may be useful summaries in technical papers, engineering manuals, internal reports, or public relations materials. Rather, most of the plant history records are in the form of operator logs, files on contract work or technical design, invoices and manifests, reports to corporate offices, and blueprints. Generally, more data are available than can be used for study purposes or digested by the investigators, but documentation could be used to test particular assumptions about plant exposures, particularly when there have been significant modifications to a plant or a change in process.

## **LIMITATIONS OF OCCUPATIONAL EPIDEMIOLOGY**

Epidemiology is often subject to criticism. Critics, in particular Alvin Feinstein and more recently Stephen Molloy, have explored issues of scientific validity in the practice of epidemiology and have disparaged the lack of scientific certainty in most observational epidemiologic studies compared to that achieved by randomized clinical trials. If it were possible to randomly assign workers to jobs and to control all aspects of their lives, the science would certainly improve. Because it is not—and few people would care to live in a society in which it were—the criticism is rather sterile. Epidemiology apart from clinical trials, like astronomy, is primarily a science of observation precisely because such control is lacking; yet within its purview are individual problems amenable to experimental and quasi-experimental approaches. There is, of course, always room for more vigorous applications of standards in science, but the limitations of epidemiology are intrinsic and seldom due to lack of effort or imagination.

## Occupational Epidemiology

Occupational epidemiology is critical to the practice of occupational medicine in many respects, including the following:

- Evidence-based medical dispute resolution (such as adjudication in workers' compensation, third-party litigation, and insurance settlements)
- Setting priorities in occupational health and safety practice
- Designing periodic health surveillance protocols
- Supporting worker education
- Identifying possible etiologic mechanisms

It is therefore important to get it right.

Epidemiologic studies of occupational disorders share many common problems. Several of the more significant are listed in Table 3.5. Exposure assessment, dealt with in Chapter 4, is a common problem in epidemiologic studies because accurate data on exposure levels in the workplace are usually lacking, especially for individual subjects. This makes it difficult to establish refined exposure-response relationships and forces the investigator to make a number of assumptions. The long latency period of many diseases, particularly many

**Table 3.5.** Common Problems in Occupational and Environmental Epidemiology

- 
- |   |
|---|
| Access to populations of interest                         |
| Statistical power   |
| Acceptable study design                                   |
| Characterizing nature and duration of exposure            |
| Exposure assessment                                       |
| Inadequacies of the death certificate as a source of data |
| Latency   |
| Confounding exposures                                     |
| Bias  |
-

cancers, makes identification of an association difficult and greatly complicates demonstration of an exposure-response relationship.

Epidemiologic studies of many occupational disorders are also difficult to perform because small numbers of subjects often result in a low statistical power. Obtaining access to suitable groups of exposed workers is often difficult and may take some time to negotiate. Such studies are difficult to interpret because of additional exposures that may occur in the industry and other risk factors involving genetics, lifestyle, and habits (for example, smoking). These are called confounders. Sources of bias may be present that result in a spurious or exaggerated association. For example, the fact that most employed workers are healthy at the time of employment makes them different from the general population, which is unscreened for health and contains within it people who are chronically ill, institutionalized, unemployed, and at high risk for illness. As a result, working populations typically have a mortality rate about 75 to 85 percent that of the general population. (See the discussion above on the "healthy worker" effect.) Without knowing this, serious errors might be made in interpreting mortality data when comparisons are made.

A related problem has been that epidemiologists tend to use outcomes and diagnostic categories that are not clinically relevant or plausible in disease risk. For example, many studies aggregate various types of leukemia into one risk estimate, although there are important differences among the various leukemias and only one—(acute myelogenous leukemia)—is known to be associated with a chemical hazard (benzene). Likewise, studies in the literature usually lump together all the lymphomas and the several types of brain cancer, although there are important distinctions in terms of cause and disease risk. They do this in order to achieve greater statistical stability (i.e., narrower confidence interval) in the risk estimate. However, the downside is that by aggregating dissimilar or different conditions, a true occupational association may be diluted by other diseases that have no relationship to the hazard or occupation under study. The net result is a false negative, or at least a falsely reduced risk estimate. On the other hand, many of these diseases are too rare to accumulate

sufficient cases to study and reliable statistics. For example, the association between myelofibrosis and benzene may never be proven with scientific certainty, because the condition is so rare, although collateral evidence strongly suggests that there is a causal association. It could be aggregated with other blood diseases or even other forms of anemia in order to yield a risk estimate with a tight confidence interval, but that would prove nothing. In such cases, an open question in the literature is probably better than forced statistics that yield a definite but erroneous result.

Epidemiologists have tended to adopt working hypotheses that concentrate on relating clinical outcomes of exposure to the hazard without rigorously examining the susceptibility of the subject exposed and other important determinants of the response. Design constraints have limited epidemiologists' ability to assess susceptibility factors in the past, and this reality has, to a degree, caused them to overlook this important determinant of response. Biomarkers of susceptibility and genomic epidemiology have helped to overcome this limitation.

A major limitation of the extant body of literature on occupational epidemiology is that the risks have changed. The workplace of 2009 generally presents hazards of a different magnitude compared with the workplace of 1949. Notwithstanding remaining problems, the workplace in developed countries has changed drastically. New processes have profoundly altered the risk profile of historical occupations. New occupations now exist in industries that, in some cases (such as biotechnology), have never been comprehensively studied or that, as in the case of semiconductor manufacture or the formulation of composite materials, change rapidly for reasons mostly unrelated to health. Even traditionally hazardous industries for which control options are limited, such as sawmill work and firefighting, have been made safer by personal protection, improved equipment, and better management practices.

The biggest problem facing occupational epidemiology, however, is obsolescence. Occupational medicine depends on a vast and sophisticated database of epidemiological information to inform

decisions and conclusions based on causation. However, the epidemiological basis for assessing risk in many, if not most, modern occupations has been rendered obsolete by technology and changes in work organization. Old studies can no longer be depended upon to guide either prevention or adjudication of compensation, and relatively few new etiological studies are being done for reasons of cost, privacy, and access. This obsolescence must be dealt with by developing new sources of information.

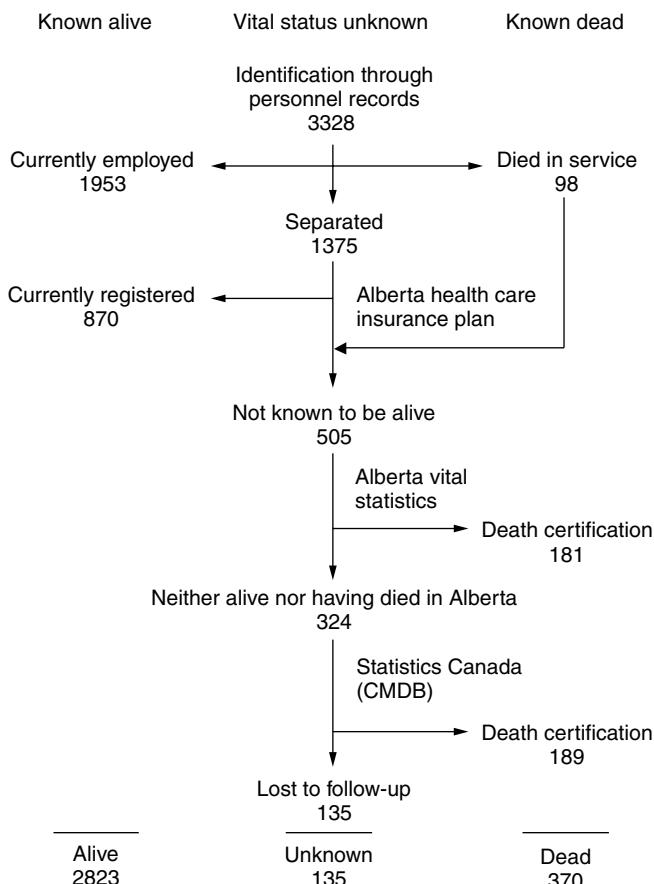
If this vast database is obsolete, the practice of occupational medicine will eventually be misguided. In particular, the assessment of causation will become further disconnected from reality. Scheduled diseases and systems based on presumption may find their assumptions to be out of date. Decisions on individual cases that are based on an interpretation of the literature will become increasingly difficult to defend.

Accompanying the profound change in the workplace environment has been a change in the working population. The general population is more fit today than in the past, and there is less difference in fitness between employed persons and the general population. The worker in a developed country today is much healthier than his or her generational counterpart of the past, due to improved living conditions and healthcare. On the other hand, many more people have entered the workforce with disabilities, and many others have stayed much longer as they age and may be undergoing treatment for medical conditions. The working population today looks more like the general population than it did twenty or thirty years ago. Occupational epidemiology does not reflect this reality.

The solution to this problem may be to revive a type of study that has long gone out of favor in occupational epidemiology: population surveillance. (See Chapter 5.)

## **ANATOMY OF AN EPIDEMIOLOGICAL STUDY**

Consider the case of a typical prospective (cohort) mortality study. Figure 3.6 demonstrates how investigators in one study set out to



**Figure 3.6.** Flow diagram for a representative cohort study: urban firefighters in the Canadian province of Alberta, 1927 to 1987.

determine the number and causes of deaths in a large occupational cohort of firefighters. Firefighters present an ongoing problem in occupational medicine, and there are many unanswered questions regarding their health risks. It is a practical issue of importance to firefighters and their families, city and county governments, fire departments in every jurisdiction, workers' compensation, and public agencies.

The investigators framed major hypotheses and concepts on the basis of preliminary data on this particular group, including that mortality from heart disease would be increased (not demonstrated, but myocardial infarction occurring on the job or just after a shift were not examined separately), that mortality from lung disease would be increased (not demonstrated), and that mortality from cancer would be increased (demonstrated for some cancers). There were secondary hypotheses as well.

How is a study like this conducted? First, one needs to identify the cohorts or group of individuals working in this occupation through personnel records, which are essential for this kind of work. The quality of personnel records is as important as that of other data sources in determining the success or failure of these studies.

The investigators reviewed the personnel files of two metropolitan fire departments to identify current workers (all of whom were assumed to be alive, of course), and separated workers who once worked for the fire department. They did not know initially whether the separated workers were alive or dead. There are several ways to find out whether an individual in a particular geographic area is alive. One of the best is to check drivers' licenses; most people hold a driver's license until quite late in life. (In Canada, it is often preferable to go through current registration for the provincial healthcare plan because health insurance coverage is universal at all ages.) This procedure removed subjects who were known to be alive and left a group whose status was "uncertain."

The investigators then linked the individual subjects with death certificates through state or provincial or federal vital statistics registries. In this example, the team first obtained death certificates from the state or provincial vital statistics bureau by going through records by name, year by year, identifying those individuals known to be deceased. They collected the death certificates and accumulated a file so that they could later study all individuals by cause of death. This also allowed them to separate the group "status uncertain" into "known deceased" and "uncertain" categories for all individuals on the original list of subjects not known to be alive. Access to a cooperative nearby vital

statistics bureau that contained data from death certificates of known quality, standardized over several years, was an enormous advantage.

For those for whom they did not find a death certificate in the province, and who may have died elsewhere, the investigators resorted to the Canadian Mortality Data Base. (The American equivalent is the National Mortality Data Base.) This automated program exists specifically for this kind of linkage. The investigators submitted a tape (this was in the 1980s) identifying all subjects with their personal identifiers, which was then processed by Statistics Canada. (The cost of accessing the Canadian Mortality Data Base was high, and it was much less expensive to get death certificates from provincial sources if they were available.) From this information, they were able to identify the cause of death for almost all subjects, leaving only a tiny number as "uncertain."

An additional arm of the study was the development of comparison data from the general population on deaths by cause by decade. At this stage, a measure of the mortality experience was calculated. The standardized mortality ratio, or SMR, was used in this study.

Table 3.3 presents the findings in two forms: as SMRs and as PMRs. As can be seen, the two measures are similar but not identical, illustrating that PMRs are simply not as good estimates of risk as SMRs and the relative risk, which is the gold standard.

An excerpt from the abstract summarizes the findings ("N.S." stands for "not significant"):

Mortality by cause of death was examined for two cohorts totalling 3,328 firefighters active from 1927 to 1987 in Edmonton and Calgary, the two major urban centers in the province of Alberta, Canada, examining associations with cohort (before and after the 1950's) and years of service weighted by exposure opportunity. The study attained 96% follow-up of vital status and over 64,983 person-years of observation yielding 370 deaths. Mortality from all causes was close to the expected standardized mortality ratio (96; 95% confidence limits (CL) 87, 107) as was that for heart disease (110; 95% CL 92, 131), and nei-

ther was statistically significant at the  $p < 0.05$  level (N.S.). Excesses were observed for all malignant neoplasms (127; 95% CL 102, 155,  $p < 0.05$ ) and for cancer of lung (142; 95% CL 91, 211, N.S.), bladder (315; 95% CL 86, 808, N.S.), kidney and ureter (414; 95% CL 166, 853,  $p < 0.05$ ), colon and rectum (161; 95% CL 88, 271, N.S.), pancreas (155; 95% CL 50, 362, N.S.) and leukemia, lymphoma, and myeloma (127; 95% CL 61, 233, N.S.); obstructive pulmonary diseases (157; 95% CL 79, 281, N.S.). Fire-related causes showed a marked excess (486; 95% CL 233, 895,  $p < 0.01$ ) but external causes overall showed a significant deficit (66; 95% CL 49, 87,  $p < 0.05$ ).

Note that the original hypotheses were not supported. On the other hand, previously unsuspected associations were discovered. As is usual in science, the study raised as many questions as it answered.

This particular study was at the time the largest study of firefighters of its type and had the most power. In the years since, other studies of firefighters have been conducted, some of them larger and with more power. The findings have been largely confirmed, although some are still controversial. The controversy, however, has opened new avenues of investigation in related fields, including issues of the evaluation of evidence and workers' compensation policies. This study is still frequently cited and has had a large impact on policies for the adjudication of workers' compensation policies for claims from firefighters. Its significance is not that it was a definitive study—those do not exist—but that it opened some new doors.

Every study has a story, and the story is usually not fully told in the paper that reports it. This study was originally undertaken to train a team of research associates in order to prepare them to do a much larger study on energy workers. It got off to a late start because a union local opposed it but the same local later agreed to allow it to proceed and in the end became its most enthusiastic supporter. The study took five years, largely because of a long delay in getting mortality data from the national database. It cost approximately \$200,000, which is extremely reasonable for a large cohort study, even at the

time. Even so, it was denounced at one point as a waste of money by a grandstanding politician on the floor of the provincial legislature. At its peak, four people were working on it at a time but usually only two. Today, such a study could be done much faster due to electronic databases, but it would probably cost at least as much.

As the OEM physician reads a study and interprets it, it is useful to think on two levels; first, on the objective level of the scientific paper. One assumes (unless there is evidence to the contrary) that the paper is accurate in its descriptions of the methods and data. What is the interpretation, given the strict construction of the science? The second is on the level of the story, which is usually untold. One may think about the “white space” in the paper—the metaphorical space between the lines and words. Why was the study done? Why was this particular group of workers chosen? What issues may have biased or driven the work and how might these influences have affected the objectivity and reliability of the work?

## **RESOURCES**

- Checkoway H, Pearce N, Kriebel D. *Research Methods in Occupational Epidemiology*. New York: Oxford University Press; 2004.
- Porta M. *A Dictionary of Epidemiology*. 5th ed. New York: Oxford University Press; 2008.
- Rothman KJ. *Epidemiology: An Introduction*. New York: Oxford University Press; 2002. [There are many fine textbooks in epidemiology, all necessarily teaching the same material. The best choice in OEM is simply the one that the physician finds most accessible and that emphasizes examples from occupational and environmental epidemiology.]
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins; 2008. [Notwithstanding the title, this is a masterful deconstruction and rethinking of the discipline of epidemiology, not a basic textbook.]

## **NOTEWORTHY READINGS**

- Abramson JH. *Making Sense of Data*. New York: Oxford University Press; 1988.

- Ahlbom A, Axelson O, Hansen ES, Hogstedt C, Jensen JU, Olsen J. Interpretation of “negative” studies in occupational epidemiology. *Scand J Work, Environ Health.* 1990;16:153–157.
- Berkman LF, Kawachi I. *Social Epidemiology*. Oxford: Oxford University Press; 2000.
- Checkoway H, Pearce NE, Crawford-Jones DJ. *Research Methods in Occupational Epidemiology*. New York: Oxford University Press; 1989.
- Feinstein AR. Scientific Standards in epidemiologic studies of the menace of daily life. *Science.* 1988;242:1257–1263.
- Flanders WD, Lin L, Pirkle JL, Caudill SP. Assessing the direction of causality in cross-sectional studies. *Am J Epidemiol.* 1992;135:926–935.
- Goldsmith JR. *Environmental Epidemiology: Epidemiological Investigation of Community Environmental Health Problems*. Boca Raton, FL: CRC Press; 1986. [A forgotten masterpiece, this book examined the use of epidemiology to address community concerns.]
- Greenland S. *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods*. Boston: Epidemiology Resources; 1987.
- Greenland S. Probability logic and probabilistic induction. *Epidemiology.* 1998;9:322–332.
- Guidotti, TL. Hidden assumptions in environmental research. *Environ Res.* 1992;59:101–113.
- Guidotti TL. Occupational epidemiology. *Occup Med (Lond).* 2000;50(2):141–145.
- Kopfler FC, Craun GF. *Environmental Epidemiology*. Chelsea, Chicago: Lewis Publishers; 1986.
- Merrill RM. *Environmental Epidemiology: Principles and Methods*. Sudbury MA: Jones and Bartlett; 2008.
- Miettinen OS. *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*. New York: John Wiley; 1985. [Classic]
- Monson RR. *Occupational Epidemiology*. End ed. Boca Raton Fl: CRC Press; 1990.
- Morris JN. *Uses of Epidemiology*. New York: Churchill Livingstone; 1975. [Classic]
- Myers JE, Thompson ML. Meta-analysis and occupational epidemiology. *Occup Med.* 1998;48(2):99–101.
- Nasca PC. *Fundamentals of Cancer Epidemiology*. Sudbury MA: Jones and Bartlett; 2008.
- National Conference on Clustering of Health Events. *Am J Epidemiol.* 1990;132(suppl 1):S1–S202.
- Neutra RR. Counterpoint from a cluster buster. *Am J Epidemiol.* 1990;132:1–8.
- Weed DL. On the use of causal criteria. *Int J Epidemiol.* 1997;26:1137–1141.
- Woolson RF, Kleinman JC. Perspectives on statistical significance testing. *Annu Rev Public Health.* 1989;10:423–428.

*This page intentionally left blank*

# **4 ASSESSMENT OF EXPOSURE**

Physicians in occupational and environmental medicine (OEM) are frequently called upon to interpret the hazard presented by a given exposure to a chemical or to correlate a clinical diagnosis with a history of exposure in either an individual patient or a group of workers. Unfortunately, accurate measurements of exposure associated with a particular patient or case are almost never available to the occupational and environmental physician in clinical practice. Physicians, in the course of their medical work, no longer perform exposure assessments themselves, although they commonly did decades ago. Even so, a working knowledge of exposure assessment is still important for patient care, understanding proposed standards, causation analysis, and assessing risk. At a policy-making level, occupational and environmental physicians may be called upon to weigh evidence for setting exposure standards in the workplace or environment based on data that cannot be interpreted without an understanding of exposure assessment.

This chapter provides an overview of exposure assessment methods from the point of view of the occupational and environmental physician, with the following objectives. The first objective is to help the occupational and environmental physician interpret the literature

on occupational exposures (both for the individual patient and for groups of workers), environmental exposures (occasionally for the individual patient but more often for populations), and exposure-response relationships. The second is to help the physician communicate with professionals from other disciplines who are more directly concerned with exposure assessment and measurement: occupational (industrial) hygienists, environmental engineers, environmental compliance professionals, and exposure assessment experts in epidemiology and risk assessment. The third is to provide some conceptual frameworks that the physician may use to think through the problem of exposure when confronted with a case in which some estimate or appreciation of exposure is important.

Data management and analysis is beyond the scope of this chapter, but the topic is presented in numerous textbooks and handbooks of occupational (industrial) hygiene and exposure assessment. Although this chapter does not discuss mathematical approaches to modeling, the reader should not be misled. The methods described are often computationally intensive and often require sophisticated database management. Since the early half and middle of the twentieth century, when occupational (industrial) hygiene became a profession, physicians have not normally been required to master or even apply these quantitative methods, unless they have been involved in epidemiological research. This chapter will therefore not attempt to present a rigorous mathematical treatment of exposure assessment.

The measurement of levels of exposure is a fundamental problem in occupational health generally for many reasons, among them:

- To determine whether levels of exposure fall within a range considered acceptable, by regulation or convention, for the protection of human health
- To determine whether workers, individually or in groups, are exposed at a level requiring that further action be taken to ensure their protection
- To identify the most likely source of an exposure in a situation in which toxicity is suspected

- To confirm that measures taken to control emissions are operating correctly
- To establish relationships between exposure and effect in research studies, particularly as the basis for evaluating present or proposed standards
- To identify which of several emissions (discharges of pollutants from the process itself) resulting in environmental pollution should be given priority
- To monitor trends, in either the workplace or general environment

Occupational exposure monitoring is similar in principle to but different in practice from environmental emissions monitoring. The technology used in the workplace is usually different from that applied to environmental quality monitoring because occupational exposures tend to be higher but confined to a more limited area. To identify emissions that might result in pollution beyond the plant boundaries, elaborate environmental monitoring systems may be required at plants or facilities, resulting in extensive collection of data at or near the plant boundary that may also be relevant to workers. Environmental monitoring is more commonly done farther away (downwind or downstream) from the suspected source, or in areas, generally close to residents, where mixing may have already occurred, in order to gain a more accurate picture of exposure in the community. All these data vary in completeness and quality, in large part because of changing regulatory requirements and particularly because of changing measurement technology.

Unfortunately, the practical applicability of environmental monitoring data to occupational health research is essentially nil. Emissions data beyond the plant boundary or from stack measurements do not reliably reflect conditions inside the plant. Winds, atmospheric chemistry, and dilution factors all modify environmental levels measured in the surrounding area. Even those monitoring stations closest to the plant may have findings quite different from background levels inside the plant, both in the chemical substances involved and

in their concentrations. Efforts to construct models that would relate environmental emissions and product throughput to fugitive emissions within the plant in an effort to derive a probable range of occupational exposures over time have not been successful, because the number of variables involved is so great and the relationship is likely to be so weak. The remainder of this chapter will specifically address in-plant, occupational exposure.

## **OCCUPATIONAL EXPOSURE ASSESSMENT**

The assessment of exposure is usually not performed by physicians, although this may change with the introduction of biological exposure indices and biological monitoring techniques. Measurement of exposure is generally performed by industrial hygienists. The technology of measurement is a large component of industrial hygiene practice and theory. Safety personnel often become involved in exposure assessment, but generally the hazards that come within the scope of their function are mechanical and physical and are recorded as present or absent.

Occupational (industrial) hygienists are the professional experts in exposure measurement, as well as in the design of controls and the solution of problems involving exposure in the workplace. Using appropriate technology, these professionals have accumulated a huge literature of area exposure-monitoring data for a variety of industries.

Although the responsibility for assessment of exposure falls primarily on the industrial hygienist, the occupational physician should be aware of certain technical aspects of exposure assessment:

- The limitations in method that might introduce an error that could place individual workers at risk
- The reproducibility of measurements and how comparable measurements may be performed by different techniques
- How measurement technology affects the setting of standards for the protection of health

- How exposure-response relationships presented in the literature are obtained
- How medical testing can provide individual exposure estimates complementary to industrial hygiene techniques
- What biomedical markers may become available soon to expand the reach of exposure assessment technology and to identify exposed workers at risk of a health effect

There are two basic approaches, or strategies, for assessing exposure in the workplace: (1) representative or typical exposure levels that reflect the usual operation of the workplace, and (2) sampling that reflects the highest exposure levels likely to be encountered under realistic conditions. The first is useful to know and essential for research on occupational hazards. The second is important for monitoring compliance, because if measured levels under near-“worst case” scenarios are within occupational exposure limits, the chances of exceeding the exposure standard under normal operating conditions are very low and there is some margin of safety.

## **THEORY OF MEASUREMENT AND SAMPLING**

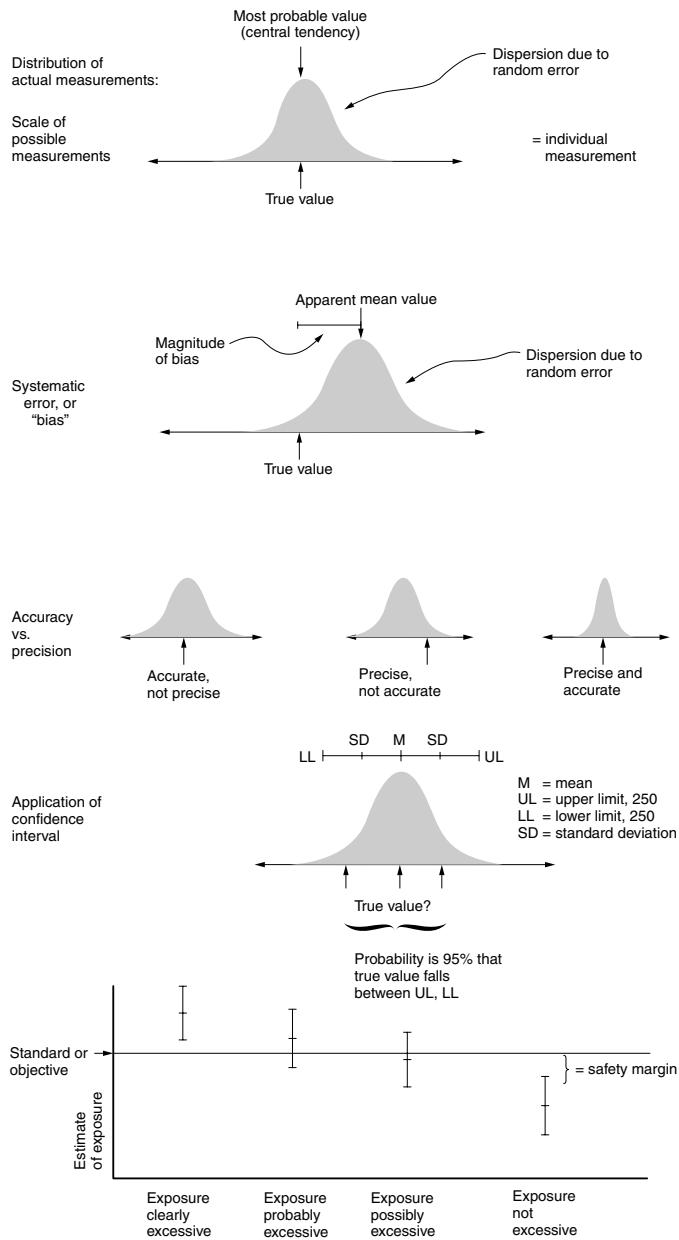
All measurements are estimates of a true value and are potentially in error. When many replicate (repeated) measurements are taken, in theory or in practice, they will cluster around a central value that should approximate the true value if the measurement system is not biased. The mean or median of multiple measurements (indicators of “central tendency”) will closely approximate the true value when a large number of measurements are taken and there is no intrinsic bias in the measurement technique. Uncertainty in the measurement is described by reference to a calculated confidence interval, usually the range of values likely to include the true value 95 percent of the time on repeated measurements. When this interval is narrow, each measurement and the mean of all measurements is a relatively close estimate of the true value, and the measurement is said to be “accurate”;

when each repeated measurement is close to other measurements, each measurement is said to be “precise,” notwithstanding that it may not be close to the true measurement. When repeated measurements on average estimate the true value closely, regardless of the dispersion of individual measurements, the estimate derived from the average is said to possess accuracy; when repeated measurements fall close together, regardless of proximity to the true value, they are said to possess precision. Figure 4.1 illustrates these definitions.

Bias arises when there is a systematic error—one intrinsic to the measurement process—that causes each measurement to have a certain probability of either over- or underestimating the true value. Measurements that are only slightly biased can be acceptably accurate; measurements that are highly biased are consistently inaccurate, although they may be very precise. Bias is usually fairly consistent, in that it remains comparable in magnitude from one measurement to another. A particular instrument, for example, usually reads consistently high or low if there is a calibration error. Extraction of a chemical from a piece of filter paper or medium is usually consistently efficient for a given solvent. If a consistent bias can be quantified, a correction factor can be applied to the data to refine the estimate to be more accurate. When bias is inconsistent—for example, when there is an alteration in the technique, spoiling of reagents, deterioration in the apparatus, or a change in the quantity to be measured from one extreme of a method’s detection accuracy to another—it can be difficult to identify and correct.

The estimate should be as accurate as possible in order to reflect the actual state of affairs, and because future measurements, possibly using different technology, should be as comparable as possible. However, it is usually more important for repeated measurements to reflect the magnitude and direction of a trend than to closely approximate the true value, especially if the true value is low compared to health standards and toxicity thresholds.

Systematic error, or bias, is generally a larger problem than random error because random error can be accounted for by statistical means. Conventional statistical treatment assumes that measurements are randomly in error and uses sampling theory to construct statistical



**Figure 4.1.** Theory of measurement and sampling as applied to occupational (industrial) hygiene.

statements describing central tendency and dispersion. The same statistical basis underlies inferential statistics as used in epidemiology (see Chapter 3). However, random error presents its own problems when accuracy is required or expected.

Monitoring for compliance, where the measurement is assumed to be reasonably accurate, must deal with the uncertainty introduced by random error. Because there is an error band, or confidence interval, around the estimate, one cannot be completely sure that the true value does not lie at or even beyond either extreme. Therefore, if the upper limit of the confidence interval overlaps or exceeds a health standard, there is a possibility that a violation may exist. It is safer for the entire interval to fall well below the standard and to have a margin of safety to ensure that violation does not occur. For this reason, many occupational exposure standards incorporate an “action level,” requiring attention at levels well below the actual standard (usually half).

Measurement techniques are frequently refined or replaced by techniques and instrumentation of increased precision and accuracy. As measurement technology advances, exposures that were previously ignored or invisible take on new significance. For example, dioxins and furans were difficult to detect below parts per billion (ppb) levels until advances in the 1980s made accurate measurement feasible down to parts per trillion (ppt) levels, and detection is not only possible but routine at parts per quadrillion (ppq) levels today. At these detection levels, dioxins and furans were suddenly found to be nearly ubiquitous. Likewise, improvements in measurement technology have demonstrated that drinking water in most American cities contains trace amounts of pharmaceuticals from sources upstream. The significance of exposure at such low levels to human health, if any, is not clear, and present control technology cannot ensure compliance to literal zero levels of exposure.

Table 4.1 lists the general approaches used for assessing exposure in occupational settings. The approaches can be classified as direct (involving direct measurement of exposure), indirect (involving measurement of an effect of exposure), or inferential (estimating without measurements specific to the exposure situation).

**Table 4.1.** Approaches to the Assessment of Exposure

Direct Assessment	Indirect Assessment	Inferential Exposure Estimates
Environmental measurement	Biological effect monitoring	Extrapolation, categorization, modeling
Personal exposure measurement	Biological markers of exposure	Simulation
Biological exposure measurement		Surrogate measurement

## DIRECT ASSESSMENT OF EXPOSURE

The technology used for direct assessment of exposure to a substance reflects the chemical or physical characteristics of the agent. A wide range of techniques is available; most of them are based on the collection of the medium, usually air, over a time period, followed by determination of the amount of the material collected and calculation of a concentration by dividing the mass of the agent of interest by total volume of the medium collected.

Direct assessment may be area-wide, at the workstation of interest, or personal, determining the individual exposure of a particular worker. The implications of these differences are discussed below in very broad terms for the physician. For technical detail and a more complete explanation, the reader should consult a textbook on occupational hygiene.

### Techniques of Occupational Hygiene

Most techniques used in occupational (industrial) hygiene for the assessment of exposure levels pertain to the measurement of concentrations of chemicals and aerosols in air. The standard method of air sampling requires a calibrated pump capable of moving air at a preset, accurately controlled velocity at either high flow rates

(about 1.0 to 4.0 l/min) or low flow rates (about 1.0 to 200 ml/min). Accurate calibration of flow is critical because calculation of the concentration depends on knowing the volume (flow multiplied by time) of air that has passed through the collection device. Aerosols are usually sampled at a high flow rate (typically 2 l/min) through a “cyclone,” which uses inertia to separate out and remove particles or to deposit them on a filter. Most organic vapors and carbon monoxide are sampled at low flow rates using a device that absorbs the compound of interest in a medium. An “impinger” is a liquid trap that captures the material in dissolved form. Activated charcoal tubes are used to adsorb many organic vapors. The entire assembly consists of an inlet (often just the exposed end of a tube, but often a fitting that protects the opening), tubing, flow-control and measurement devices, a collection device, and a pump—this is called the “sampling train.”

Aerosols can be quantified as to size and number or mass by several means. Weighing is basic. Direct visual counting by microscopy is tedious and does not give information on the distribution of mass by size, but it is necessary for fibers, such as asbestos, or to identify silica. Direct inspection also allows qualitative sorting, such as determining different types of asbestos in a given sample or separating out asbestos from other fibers such as talc, wood, or cotton. In most circumstances, however, the key determination is the difference between total aerosol concentration in the air and that of “respirable particulates” (less than 10  $\mu\text{m}$ —see Chapter 2). This can be accomplished quickly by the procedure known as “two-stage sampling,” in which a cyclone is placed in the sampling train in front of a fine filter. The cyclone separates out all particles 10  $\mu\text{m}$  and greater in diameter; the filter captures particles less than 10  $\mu\text{m}$  in diameter. The weight of the particles deposited on the filter divided by the total volume of air passing through the filter yields the mass concentration in  $\text{g}/\text{m}^3$ . The weight of particles deposited on the filter and separated by the cyclone, divided by volume, yields the total aerosol concentration in the same terms. The dust deposited on the filter is also often examined by X-ray diffraction or infrared spectrophotometry to determine the silica content.

Gas and vapor sampling in the traditional manner involves capturing the material in an impinger or adsorption tube and then analyzing the media to quantify the amount present; divided by the volume of air passing through the sampling train, this yields the concentration. The technique may involve “wet chemistry” at the laboratory bench, gas chromatography, mass spectrometry, or other techniques, depending on the chemical reactivity of the material. In the case of detector tubes, this same idea is streamlined into a more qualitative colorimetric test; accuracy is sacrificed for quick results in the field.

For rapid determination of exposure levels to fumes and gases in the field, “indicator” or “detection” tubes are often used. These consist of transparent tubes of plastic or glass containing a sorbent, which traps the material and is coated or impregnated with an indicator chemical that changes color in the presence of the chemical being tested for. When a fixed volume of air is passed through the tube by use of a simple hand-operated pump, a change in color indicates the presence and, within limits, the magnitude of the concentration of the chemical hazard.

In recent years “passive dosimeters,” collection devices not requiring pumps, have improved greatly in collection efficiency and reliability and are replacing active collection systems for many applications. The device consists of a permeable membrane over a sorbent, such as activated charcoal, packaged in the form of a small badge. These can be worn by workers for convenient personal-exposure monitoring. The chemical is collected by diffusion across the membrane and by adsorption on the sorbent material, which is then washed with a solvent that extracts the bound chemical. The “solute,” thereby extracted, is then analyzed by gas chromatography. Passive dosimetry is increasing in technical sophistication, and the variety of chemicals that can be assessed in this manner is increasing.

Direct-reading meters have been developed for a wide range of chemical and physical hazards based on unique characteristics of the hazard, such as peaks of infrared absorption in the case of carbon monoxide detectors. The most popular direct-reading meters are those that test for oxygen (“oximeter”); carbon monoxide (“carboximeter”);

explosive hydrocarbons (“explosimeter”); and physical exposures such as heat, light, microwave radiation, and noise. Noise monitoring is somewhat more complicated and may involve several different instruments, such as sound-level meters (for instantaneous readings of continuous noise, as loudness by frequency), dosimeters (for time-arrayed readings, as arrayed loudness by frequency), and impulse meters (for the sound-wave pressures generated by sudden, pulsatile noise) (see Chapter 9).

The design of a specific sampling strategy is a challenging task, requiring knowledge of the plant and the timing of the process and workday. Under certain circumstances, it may be adequate to take isolated, arbitrarily selected “grab” samples to determine peak exposures or those occurring as the result of an accident or in a confined space. Ordinarily, however, sampling should take place over time in order to determine peak and average exposures, with particular reference to those times and locations when exposures are maximal. Assumptions on exposure should be confirmed by area monitoring. Occasionally, materials are tracked much more widely in a plant than is realized, or environmental background levels may be higher than is supposed. Even office workplaces should be monitored occasionally for airflow, lighting, noise, and whatever chemical exposures may be nearby or possibly intruding through communicating spaces or shared ventilation systems.

In designing an industrial hygiene survey, it is important to decide which situation is more important to identify: the worst case that may exist in the workplace or the most representative situation in the workplace. If the survey is being conducted as part of a long-term study or to assess the hazard in a new process, more typical or representative locations and times may be selected. When the survey is conducted for purposes of assessing compliance with regulations or to evaluate a worst-case scenario, it is most useful to pick those locations in which exposure is likely to be highest and to measure the level of exposure at periods when work intensity or use of the chemical in question is highest. This will provide an estimate of the maximum exposure likely to be encountered.

Therefore, monitoring for compliance is, or should be, a search for the situations that present the greatest risk to workers. On the other hand, if the objective is to conceal adverse working conditions, measurements can be made at places and periods of low activity, producing misleadingly low results. It is not acceptable, but it is unfortunately common that a sampling strategy may be designed to intentionally minimize estimates of exposure. It is important that the occupational and environmental physician be aware of and alert to this possibility. In recent years, forensic hygiene methods have become increasingly important for detecting such unethical practices and supporting their prosecution.

Measurements are often undertaken as part of the investigation of a particular problem, and so incident investigations tend to reflect unusually bad exposure situations. These investigations are called “health hazard evaluations” (HHEs) by the National Institute of Occupational Safety and Health (NIOSH). Health hazard evaluations are invaluable for assessing exposure to particular hazards but are usually limited to one or a small number of hazards. They are based on an identified health problem, employee complaints, or the observation of occupational health professionals conducting periodic inspections.

Another key element in exposure assessment strategy is the decision to employ area monitoring or personal monitoring. Area monitoring estimates the exposure in a given area or at a given work station. If workers move around, or if their duties involve multiple sources of exposure, this may not accurately reflect their actual exposure experience. In such cases, area monitoring may yield misleading results. The preferred approach then is personal monitoring.

Personal monitoring involves the collection of data on individual exposures by portable continuous sampling devices that clip onto the worker’s lapel, belt, or other clothing. In personal monitoring of airborne hazards, the inlet should be in the worker’s breathing zone; the easiest way to ensure this is to clip the end of the tube to the worker’s lapel or the front of the worker’s shirt. Personal sampling provides much more specific information on exposure and is usually

used in conjunction with area monitoring for a more complete profile of exposure possibilities. In each case, individual workers must be informed of the results of measurements relating to their own areas or person, and copies of the reports including the actual measurements should be placed in each worker's medical record for future reference.

NIOSH publishes standard methods for occupational hygiene exposure measurements. Deviations from these standard methods should only occur for good reason, and new or nonstandard methods must be documented and validated against the corresponding NIOSH standard methods.

### **Biological Exposure Measurements**

Another category of direct measurements of exposure is based on the detection of the chemical or its metabolites (see Chapter 2) in the body fluids (usually urine) or exhaled air of the exposed workers. These biological exposure measurements reflect the internal dose or body burden absorbed by the individual worker and are thus complementary to the personal exposure monitoring described above. Biological exposure measurements reflect actual individual absorption from all sources; personal exposure monitoring reflects the potential for individual absorption. Biological exposure measurements are able to present a more complete picture of total absorption, although they cannot distinguish among exposures from different sources. Standards based on biological exposure measurements have been developed, and those that are recommended by the American Conference of Governmental Industrial Hygienists (ACGIH) are called "biological exposure indices" (BEIs). BEIs are currently replacing many traditional exposure standards because of their ease of measurement and reliability.

Biological exposure measurements carry a number of limitations. There is relatively wide variation in individual handling of absorbed chemicals, so it is not usually possible to extrapolate from biological exposure measurements to personal or area exposure levels with accuracy. Not all chemicals of occupational importance lend

themselves to this method, because of limitations in practical measurement technology, chemical reactivity, or lack of mobilization after initial absorption and distribution. The approach works particularly well for organics, especially volatile hydrocarbons, lead, and carbon monoxide. It works much less well for most chlorinated polycyclic hydrocarbons (except pentachlorophenol) and most potentially toxic gases. Timing of the collection is particularly important for most chemicals because of metabolism and rates of excretion. When applied to urine, measured concentrations must be corrected by normalization to the creatinine concentration, to account for dilution.

Examples of biological exposure measurements that are now well established include carbon monoxide in end-expired air taken at the end of a shift (to reflect maximum carboxyhemoglobin levels), lead levels in the blood, styrene in mixed expired air before and after a shift, toluene or benzene in end-expired air, and total phenol in urine at the end of a shift (reflecting phenol or benzene exposure). ACGIH is steadily adding to the list of recognized methods, and these measurements have become part of the standard practice of industrial hygiene and occupational medicine in some industries. Because analytical methods for these measurements are very accurate and sensitive, low exposure levels can readily be documented. Individuals can be observed over time and studied repeatedly without inconvenience. The cost of this technology is rapidly falling, and these tests are becoming generally available at relatively low cost compared to area or personal exposure assessment methods. Biological exposure measurements based on concentrations of xenobiotics or metabolites in urine or expired air are certain to play a much larger role in the future in influencing the practice of occupational medicine. They also represent an area of overlap between the responsibilities of physicians and those of the industrial hygienists. Occupational physicians generally do not consider this to be an encroachment on the practice of medicine. Although they are tests on people, these tests are primarily for the purpose of assessing exposure, rather than for diagnosis or for assessing health status. These tests are generally conducted by hygienists unless it is necessary to draw blood or to conduct invasive procedures.

## INDIRECT MEASUREMENT

Often, it is not possible to assess exposures directly. This is a fundamental problem in methodology in epidemiological studies in which exposures went unmeasured or were incompletely measured in the past; in which workplaces are inaccessible or access is not allowed; in which exposure is immeasurable for technical reasons; in which exposure occurred in the distant past, before technology was available for measurement; in jobs in which sampling was never performed; involving avocational exposures; involving multiple exposures (particularly in common cumulative exposures such as lead); involving disputed cases in which information is suppressed; and in which a hazard has not previously been suspected. Certain strategies are available to estimate exposure and to reduce guesswork in such situations:

- Biological monitoring, which may approach direct measurement in accuracy
- Biological markers
- Extrapolation, categorization, and modeling
- Simulation
- Surrogate measurements

The first two are indirect means of assessing individual exposure. The last three are ways of making logical inferences about what levels of exposure might have been for groups of individuals. Each is described in more detail below.

### Biological Effect Monitoring

Biological effect monitoring is the measurement of the effect of an exposure on a physiological response in the body or a critical biochemical component that is not involved as the metabolite of the agent. In the past, the most familiar examples of biological effect monitoring were the “ZPP” (zinc protoporphyrin) and “FPP”

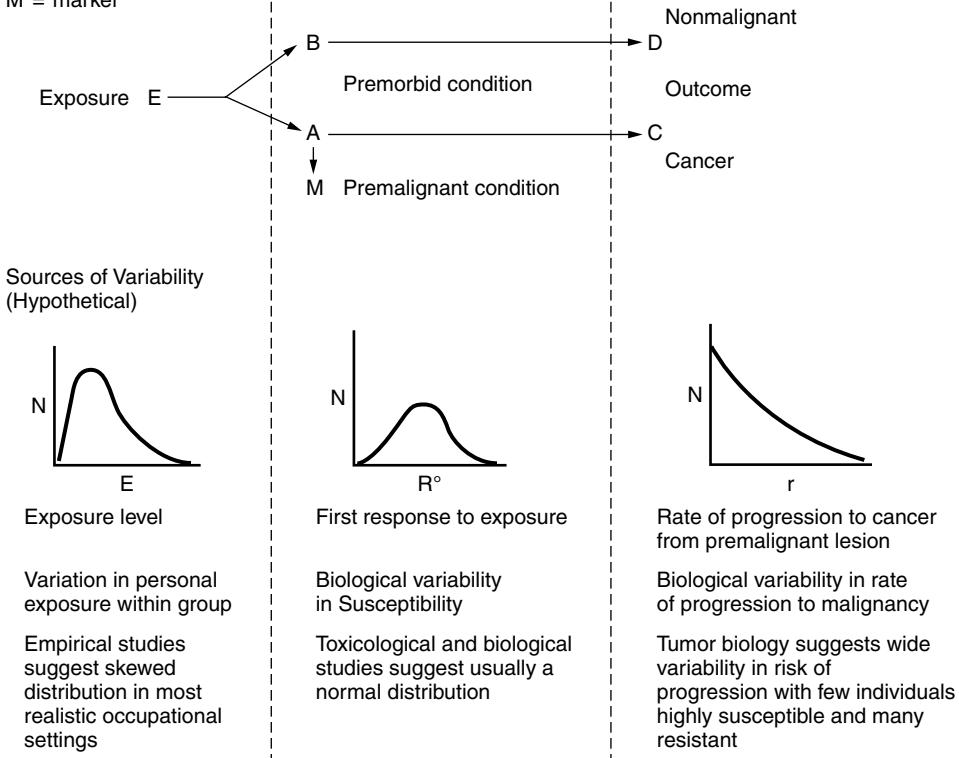
(free protoporphyrin) assays for the degree of inhibition by lead of an important synthetic enzyme in the heme pathway, but these methods have fallen into disuse because their thresholds were too high. Today, the most familiar example is probably the plasma or red-cell cholinesterase assay for the effect of organophosphate pesticide exposure (see Chapter 10). In both examples, the effect reflects the internal dose of the agent. There is some degree of individual variation in both because of differences in baseline activity and susceptibility to the effects. Biological effect-monitoring tests have been incorporated into the BEIs proposed by the ACGIH. The first to be accepted was the ZPP, although it is no longer used.

Biomarkers are discussed in detail in Chapter 2. Biomarkers of effect may include indicators of mechanism or activity at the cellular, genomic, or molecular levels. These are more commonly used in research. Adducts of genotoxic molecules that are covalently bound to DNA, or dimers of abnormally bound twin DNA bases resulting from excision and repair, can be detected in urine by sophisticated techniques, as can adducts to other macromolecules. In the future, these tests may become practical monitoring techniques in some applications.

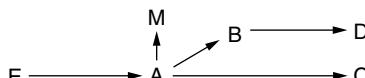
Some biological markers are clinical. For example, the unusual skin condition known as chloracne is a marker of exposure to organochlorine compounds, and palmar keratoses are markers of exposure to arsenic and therefore reflect cancer risk indirectly. Figure 4.2 outlines the use of clinical markers as surrogate indicators of exposure for later disease, using cancer as an example, and shows the many different relationships that may exist between the marker and the underlying pathway of disease, the biological variability in susceptibility, and the resulting uncertainty in outcome.

Unfortunately, there has been considerable confusion over terminology, and a useful distinction has been lost. The term “biological monitoring” was originally defined as biological effect monitoring. However, the term “biological monitoring” has been applied in recent years to biological exposure measurements. The original distinction, implying only indirect assessment of exposure, would

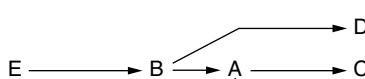
- A = premorbid process resulting in marker  
 B = premorbid condition not directly related to carcinogenesis  
 C = carcinogenesis  
 D = nonmalignant death outcome  
 E = exposure  
 M = marker



Alternative relationships



Premorbid condition may relate to carcinogenesis process



Premorbid condition may be an intermediate step in carcinogenesis process



Premorbid condition may be unrelated to carcinogenesis



**Figure 4.2.** Schema for understanding the role of biological markers. Reprinted from Guidotti, T. L. Hidden assumptions in environmental research. *Environmental Research*, 1992, 59(1): 101–13.

have been a useful one to maintain, but it eroded quickly. To preserve the definition, the rather ungainly and nonstandard expression “biological effect monitoring” has been used here.

### **Extrapolation, Categorization, and Modeling**

This group of techniques is based on reconstruction of the most probable exposure level on the basis of data not specific to the group or individual being assessed. The exposure can be estimated on the basis of the best available information, but usually the probable levels of exposure are classified in broad rank categories such as “nil,” “low,” “medium,” or “high” and often just “unexposed” or “exposed.” Their principal application is in epidemiological research.

Exposure information can sometimes be inferred from plant records or extrapolated from levels in other areas of a plant or from similar plants. In many industries, plant engineering records are kept at both the corporate and plant level and can be used to reconstruct probable exposure occurring in the past. Government agencies with regulatory authority over the industry may also have supplemental plant information available, which, because of standardized reporting requirements, may be in a form more easily accessible and interpretable by outsiders. Unless the agency has the mandate for occupational health protection, however, these records would not explicitly describe in-plant exposures and would contain mostly information irrelevant to a proposed study. Government offices may also have some plant information available relating to particular incidents, such as accidents resulting from environmental releases or fatalities. These can all be used to make “best-guess” estimates, but they require interpretation by an experienced exposure assessment professional.

The most accurate application of extrapolation comes when data on one type of exposure can be reliably extrapolated to suggest the level of exposure to another, related exposure. For example, in one study information was available on total suspended dust particles in a mine over several years, but only recent information was available for dust levels in the respirable range. The two, however, are related. The estimation of

exposure to respirable dust over time was possible using good, recent data on the fraction of total dust that was in the respirable range. First, the data available from a cross-sectional study on measurements of respirable dust and total dust taken at the same time was examined. The investigators then determined the ratio of respirable total dust from the best paired data at several points in time. (For example, if the cross-sectional study was conducted over a period of a few months or a year, there may have been measurements taken in the same areas at different times.) If the ratios were approximately the same, the investigators took the average of the ratios from the cross-sectional study and used this ratio to estimate the respirable dust levels from the total dust measurements from past years as the best estimate of respirable dust over the years. This approach has worked well in a number of studies.

The extrapolation approach would not work if there had been a change in mining techniques and the technology resulted in different particle-size distribution. This is unlikely unless conditions in the mine changed dramatically. Usually when that happens, the particle-size distribution shifts to favor smaller particles and *more* dust is produced in the respirable range, because modern technology creates finer particles due to higher speed and more-forceful cutting and drilling. Plant engineering records are most useful as an adjunct resource for checking details of the plant history or confirming dates of major changes recalled in more-general descriptions of plant operation reported in internal memoranda, technical papers, or by the recollections of old-timers among the employees. The latter source is invaluable for obtaining a general orientation, identifying major developments among the mass of engineering detail, and recalling unusual events.

It is often a critical assumption in a proposed study that exposure levels in the plant remained the same or changed in a predictable way that could be estimated from the technical modifications. This assumption should be tested—perhaps by applying whatever serial data on exposures may exist, or by comparing monitoring data before and after technical modifications of the type performed in the past. It is a major advantage when the technology is mature in an industry. The use of personal protection and controls on hazards, such

as ventilation and dust suppression techniques, obviously changes individual exposure regardless of the characteristics of the source.

Another approach is to use theoretical models to predict the most likely exposure level for a given subject. This technique is more commonly used in environmental studies than in occupational health, and is usually restricted to water or atmospheric dispersion models for studying waterborne or air pollution health effects. Without empirical validation of the model in some way, the theoretical exposure may or may not come close to an accurate prediction of individual exposure, but the discrepancy may never be detected. The use of a model is not very convincing unless there is good evidence that the model fits reality and can be shown to predict the actual exposure profile. The usual way to calibrate a model is to see if it predicts levels of exposure consistent with historical data, if available, and to test the model against contemporary measured results.

In occupational health studies, an increasingly popular means of categorizing individuals by level of exposure is the “job exposure matrix.” This is a table or computerized program relating a worker’s job description, specific to the industry, to level of exposure on the basis of representative industrial hygiene surveys and the opinion of experienced hygienists. The matrix is actually a sophisticated table of the most-likely exposure estimates for well-defined jobs. It can only substitute in general terms for an assessment of individual workers, because individual jobs and exposure situations may differ; but the method works well in population studies, when what is important is a mean level or a rank order. It provides an estimate that corresponds roughly to the most probable estimate for that subpopulation of exposed workers, and therefore minimizes error when used for large groups.

## **Simulation**

Simulation of the process and setting can be used to provide empirical evidence on which to base an estimate of exposure. The method is used to assess exposure in workplaces that no longer exist but that can be reconstructed using period-specific technology. At its most

refined, simulation may allow the generation of a set of most-likely exposures specific to a number of job classifications and related to specific work conditions and production levels.

The concept is to reconstruct the circumstances of exposure in a controlled setting as authentically as possible, reproducing as accurately as possible the operating conditions found in practice. For example, in a major study of silica exposure in the stonecutting industry, a granite shed in Vermont was rebuilt and operated using established work methods so that industrial hygiene measurements could be taken. These measurements were used to estimate the exposure of workers many years before, at a time before measurements were available.

Simulation is only practical when the scope of work is limited and considerable information exists on customary work practices of the time. It would not be practical for studies of a large or complex plant that has closed down or changed processes with the use of new equipment. On a laboratory scale, however, some processes may lend themselves to mock-up and trial runs. Such scaled-down simulations can provide valuable information on potential levels of exposure, but in terms of reliability, they fall far short of empirical data on exposures in full-scale operations.

### **Surrogate Measurements**

Surrogate measurements are used when data exist for one chemical or component of a hazard but not for the exposure of interest. If the two are closely related, monitoring or estimating exposure to the surrogate provides an acceptable determination of the level of exposure of the exposure of interest, even though the actual exposure cannot be quantified accurately. In the example given above for the study of exposures to particulates in a mine, the investigators were trying to establish an accurate exposure-response relationship for the effects of exposure to silica and needed to estimate actual exposure to silica. They did so by deriving an estimate of exposure to respiratory dust and used that as a surrogate measurement for silica exposure, because they could not measure silica exposure directly.

The most common use of surrogate measurements is probably the benzo(a)pyrene measurements to indicate levels of exposure to polycyclic aromatic hydrocarbons (PAHs) from combustion products following fires, diesel or auto exhaust, and cigarette smoke. The relationship between benzo(a)pyrene concentrations and concentrations of other hydrocarbons of interest is not usually accurately known and varies from situation to situation. In all cases, however, levels of exposure to other hydrocarbons tend to move in parallel to benzo(a)pyrene when they are driven by a common source. Because it is relatively easy to measure and is always included on constituent hydrocarbon analyses, benzo(a)pyrene serves as a useful surrogate for other PAHs that might contribute to toxicity.

### **Markers and Biomarkers**

Markers of disease are signs and symptoms that are used clinically to assess disease risk. Some clinical markers of exposure are disorders themselves. Although few in number, these clinical biomarkers are particularly important for the clinician to recognize. Table 4.2 lists various clinical markers now recognized as indicative of exposure

**Table 4.2.** Clinical Conditions That Serve as Markers of Exposure and Risk of Cancer

Clinical Finding	Cancer Site	Exposure
Palmar keratoses	Bladder, lung	Arsenic
Airway obstruction	Lung	Cigarette smoke
Leukoplakia	Oral cavity	Cigarette smoke
Chloracne	Soft tissue (sarcoma)	Dioxin, phenoxy herbicides
Pleural plaques	Lung cancer, mesothelioma	Asbestos
Skin erythema, photosensitization, hyperpigmentation	Skin, lung	Polycyclic aromatic hydrocarbons

and suggesting possible future risk. The best known is probably chloracne, the characteristic dermatitis associated with exposure to chlorinated polycyclic hydrocarbons (see Chapters 10 and 17). In studies investigating exposure to the PCBs or dioxins, chloracne has been used as a clinical marker suggesting exposure. Ideally, a marker should reflect past exposure and also suggest future risk. This is clearly the case for palmar keratoses associated with arsenic exposure, but as it happens the presence of chloracne does not predict the risk of future malignancy. Pleural plaques can identify a subject as belonging to a group in which asbestos exposure has occurred, but their presence or absence in a person known to have sustained exposure does not add additional information predictive of future risk.

Biomarkers are signs or (more frequently) field or laboratory tests that reflect what is happening in the body with respect to exposure (which is of interest in this chapter), susceptibility, effect, and disease risk (which are discussed in Chapters 2 and 10). The concept and theory of biomarkers have been primarily driven by epidemiology and the need for toxicological indicators on a population basis. Their use in clinical medicine is more familiar as screening tests for prevention and early detection (secondary prevention) and for diagnostic testing. (See Chapter 2 for a more complete discussion of biomarkers.)

A biomarker of exposure is defined by the National Research Council as “an exogenous substance or its metabolite(s) or the product of an interaction between a xenobiotic agent and some target.”

The leading organization developing practical biomonitoring guidelines for the workplace is the American Conference of Governmental Industrial Hygienists (ACGIH). ACGIH and its role in developing the guidelines known as threshold limit values (TLVs), a voluntary standard for the workplace environment, are discussed in Chapter 7. The ACGIH-recommended counterpart guideline for biological monitoring is the Biological Exposure Index (BEI), which is used as a voluntary standard for individual exposure. A BEI is a biomarker of exposure, which means that it is a guideline applied to a measurement that reflects internal dosage, not environmental

exposure. BEIs have been developed for a variety of potential exposures, based on knowledge of toxicokinetics, practical measurement technology, and assessment of intra- and inter-individual variation for purposes of standardization. Workers may be exposed to the same chemical (such as solvents) at work and at home, as well as in avocational pursuits such as hobbies or volunteer work.

The strategy for taking measurements (usually post-shift) ensures that occupational exposure dominates in the event of multiple sources. The timing depends on the toxicokinetics of the agent of concern (see Chapter 2). Volatile organic compounds should be measured in expired air immediately post-shift. Blood-lead levels represent cumulative, not recent, exposure and should be monitored periodically and over a period of weeks, not days. Serum cholinesterase levels for organophosphate exposure respond quickly, but the exposure is intermittent—not tied to a shift—and so should be measured at baseline, before any exposure occurs, and then after significant exposure and in the event of symptoms. For reasons of compliance, occupational justice, and legal requirements, sample collection for BEIs is done on the employer's time, not on the worker's time after the workday has ended.

## **A PRACTICAL APPROACH FOR THE CLINICIAN**

Historically, occupational hygiene was part of occupational medicine, and physicians made the exposure measurements—especially in the days of Legge, when “factory inspectors” were physicians. Today, occupational hygiene is its own discipline, and hygienists and exposure-assessment experts are trained and equipped to make exposure measurements better than most OEM physicians. However, the OEM physician needs to be able to think about exposure in subtle ways and to relate exposure patterns to the risk of disease in the individual subject. This is very helpful in clinical work and essential to causation analysis.

The translation of the toxicological exposure-response relationship into an epidemiological exposure-frequency relationship is complicated

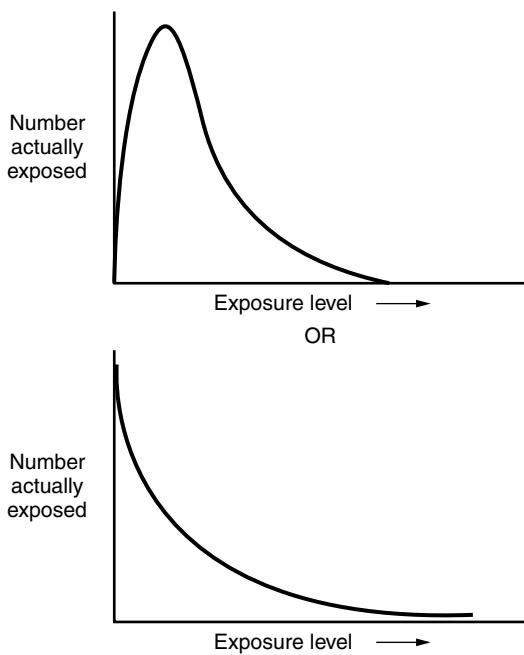
by uncertainty in the estimation of individual exposure. The considerations of sensitivity, accuracy, and precision common to all measurement techniques apply to exposure quantification as well as to clinical tests (see Chapter 5). As a practical matter, the technology for measuring exposures directly is not a limiting factor for most purposes. Rather, the difficulty lies in estimating exposure for an individual. Except for organochlorine compounds and some "heavy" metals, relatively few xenobiotics accumulate or persist in the human body long enough to be measured accurately months or years after the fact. Thus, unless the exposed worker or patient is encountered in the emergency room or soon after, the clinician, particularly in a referral practice, has little chance of quantifying the level of exposure with any degree of certainty.

Individual measurements of exposure in the workplace over time are rarely available, except among workers exposed to radiation, where the use of film badges is almost universal. Instead, our best estimates come from personal or area sampling. This is seldom performed systematically for an individual subject. More commonly, only area measurements are available. Most commonly in clinical practice, there are no exposure data at all, or records are so fragmentary that they can be used only to establish a probable range of exposure.

Although individual circumstances differ and a variety of factors may distort the pattern, some generalizations can be made on theoretical grounds, which fit empirical data reasonably well. Where hazardous exposures exist in the modern workplace, they are usually contained or controlled to some degree, and contact with them is limited by personal protection of workers up close. Thus, exposure right at the source of a fugitive emission is unlikely, except in unusual or upset accidents or outright negligence.

As an emission escapes, it is subject to dilution in the medium (usually air) within the space of the plant or workplace. Air currents will accelerate mixing; physical obstacles to diffusion will delay it. A single release diffuses into the volume of the plant, but a continuous release will fall in concentration, all other factors being equal, by at least the inverse square of the distance from the source, if there is

mixing. If indoors or if the agent is heavier than air, as with hydrogen sulfide, or is released into a confined space, this decay in concentration is delayed because of inadequate mixing. Diffusion occurs more readily outdoors, where the height of diffusion is not restricted, or if the agent is lighter than air. Thus, the idealized curve for the probability that a given individual has been exposed in a realistic workplace setting is roughly that shown in Figure 4.3. Virtually nobody within the plant boundary would have no exposure whatsoever, however minimal, to a volatile agent, but few will sustain very high exposures. In a workplace with a random distribution of workers on the shop floor, or when there is a random distribution of the time spent at dispersed locations by workers and a single point-source of fugitive emissions, a larger number of workers will, predictably, sustain lower exposures. The shape of the curve usually assumes a logarithmic



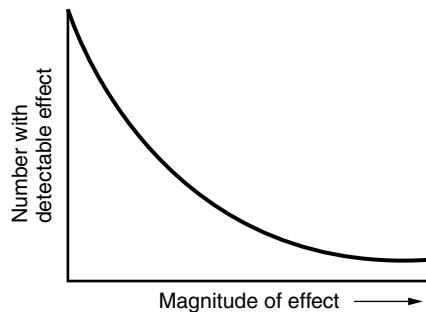
**Figure 4.3.** The opportunity for exposure in the workplace in realistic situations is usually skewed; most subjects are exposed to lower levels, and relatively few are exposed to high levels.

distribution and as the distance from the source increases, the concentration falls, but the number potentially exposed rises. Obviously, few workplaces have uniform or random distributions of workers in relation to exposure to a single source, but in operations where workers move around from site to site frequently, as in a natural-gas plant, and the most hazardous operations are contained and monitored, this model is not a bad first approximation.

Uncontrolled, or “upset,” releases may result in a small number of workers sustaining much higher exposures for short periods. Where the toxicity of the agent is cumulative, as in lead, the additional, higher exposure must be compared to the much longer opportunity to absorb the agent at low levels. Unless massive, the time-weighted average exposure over a shift may still be relatively modest. Very high uncontrolled exposures, sufficient in themselves to produce toxic effects, are not common, but devastating when they occur. They are subject to the general principle in safety that the probability of an incident occurring is inversely proportional to its severity, in large part because that is precisely how engineers and millwrights design things.

When a short-term, high-level exposure occurs, it is likely to be qualitatively different in its effects from lower-level, longer-term occupational exposures, despite simplifying assumptions that the concentration-time product is a constant for toxicity (see Chapter 2). Conventional occupational exposure standards use a time-weighted average on the assumption that the concentration and the duration of exposure are equally important in determining toxicity. This is rarely true. For many gases (for example, hydrogen sulfide), concentration is much more important, and acute, short-term exposures may be much more toxic than longer-term, lower-level exposures. The time-weighted average is acceptable toxicologically only because occupational exposure levels are generally well below the toxicity threshold (although they do not protect every worker).

Because the majority of individuals in an exposed population are likely to be exposed at relatively lower levels, and because only a fraction of the population is likely to be susceptible to the effects that can be measured clinically, only a few individuals are likely to show



**Figure 4.4.** In realistic exposure situations, usually few subjects show serious effects, and most show modest or no detectable effects.

major effects in most realistic exposure situations (see Figure 4.4). This is why even catastrophic events usually involve relatively few deaths or serious casualties compared to the much larger population that is exposed in the event.

## EXPOSURE ASSESSMENT AND HEALTH-RISK ASSESSMENT

Personal exposure monitoring is by far the best environmental exposure data if the objective is to pair individual exposure levels with individual risk. Personal exposure is usually representative of the exposures in a particular job and may detect incidental exposures associated with work practices that would not be reflected in the workplace generally. For example, the use of solvents to clean equipment or the maintenance of filters, traps, or catalyst cartridges may result in significant exposures not reflected in occupational hygiene reports. The next best thing would be comprehensive occupational hygiene surveys. Area monitoring is used to derive an estimate of the exposure range in the specific workplace, but not for specific workers, who may move among many locations and may perform tasks involving additional exposures that would not be reflected in area levels.

Both health hazard evaluation and routine area monitoring have the additional disadvantage of normally being designed to detect the

highest exposure levels that occur, rather than representative levels. Thus, the selection of monitoring apparatus (often a threshold-alarm device or a semi-quantitative measurement such as a colorimetric Drager tube), the location of measurements (usually in areas of the workplace felt to be most hazardous or to present the greatest exposure), the timing of the measurements during the work shifts (if not continuous), and the selection of what exposures are to be measured are all intended to identify the maximum likely exposure and to ensure that it is below the occupational exposure level (OEL) or a predetermined threshold limit value (TLV). This is sound occupational hygiene practice because operational exposure monitoring is intended above all to protect the health of the worker. It is not ideal for occupational health research, however, because the association of a health outcome with exposure levels requires representative exposure measurements whenever possible, both to identify associations below the OEL and to construct exposure estimates for individual workers so that they may be accurately categorized, and so that those with higher exposures can be separated from the others. To avoid falsely attributing health outcomes to a single exposure, it is also desirable to have similar exposure measurements for all other exposures likely to be significant, some of which may be confined to a small number of workers within the larger group.

In research, data from health hazard evaluations and from routine area monitoring are useful primarily for two purposes: (1) to guide the design of a more comprehensive exposure monitoring study when this can be done, and (2) to identify trends in exposure over time when changes in plant layout or technology modify the exposure patterns in particular workplaces. These data can be used to estimate probable exposure levels when they are the only available measurements, but the uncertainty of the estimate will be greater, and fine distinctions among job classifications or work stations may not be possible.

Sometimes, the only occupational exposure monitoring that has been done is by the use of alarms in plant facilities and by portable meters in the field. When available, these data provide an indication

of the frequency of exposure but not necessarily the magnitude or level. For example, in natural gas facilities (such as pumping stations), the technology of hydrogen sulfide monitors in the 1980s was such that an alarm might sound if concentrations rose above a certain threshold, but accurate measurements were not possible and there were many false alarms; sometimes workers disabled them because they were so annoying. Because the purpose of the measurements was to protect the workers from acute effects and not to identify events of exposure, no reason existed at the time for systematically recording the measurements. Alarm data are seldom recorded in ways that would allow construction of an exposure profile for a particular plant or workplace.

## RESOURCES

The American Industrial Hygiene Association (AIHA) publishes a wide variety of materials on exposure assessment in the workplace. See their Web site: [www.aiha.org](http://www.aiha.org).

Nieuwenhuijsen MJ. *Exposure Assessment in Occupational and Environmental Epidemiology*. New York: Oxford University Press; 2003.

Plog BA, Quilan PJ. *Fundamentals of Industrial Hygiene*. Washington, DC: National Safety Council; 2001.

## NOTEWORTHY READINGS

Guidotti TL. Hidden assumptions in environmental research. *Environ Res*. 1992; 59:101–113.

Hawkins NC, Jaycock MA, Lynch J. A rationale and framework for establishing the quality of human exposure assessment. *Am Ind Hyg Assoc J*. 1992;53:34–41.

Schulte PA. Methodologic issues in the use of biologic markers in epidemiologic research. *Am J Epidemiol*. 1987;126:1006–1016.

Smith TJ. Exposure assessment for occupational epidemiology. *Am J Ind Med*. 1987;12:249–268.

*This page intentionally left blank*

# **5 MONITORING, SURVEILLANCE, AND SCREENING**

One strategy for protecting the health of workers is to examine them for early signs of disease. Such programs may be instituted for community residents at high risk following an unusually potent environmental exposure, but in general there is no counterpart in environmental medicine to the monitoring, surveillance, and screening programs that have been an integral part of occupational medicine practice. Monitoring is a general term for following the health experience of a worker or group. Surveillance means to follow them for a specific outcome. Screening is testing to detect a disease or condition in a person who does not have symptoms suggesting the disorder.

Monitoring, surveillance, and screening are among the oldest services provided by practitioners in occupational medicine, but they have fallen out of favor in contemporary practice. There is probably no type of service in occupational medicine that has contributed more to the development of the field or that, at the same time, has detracted more from its credibility. This paradox arises from the history of the field and trends in the health of the workforce. Screening tests (and pre-employment examinations, as they were called at the time) were introduced well over a century ago, in large part to identify workers who had clinical diseases that would render them unfit

to work and pose a risk to fellow workers. The disease of greatest concern, of course, was tuberculosis, which could be transmitted in the workplace under crowded conditions. In addition, the results of induction examinations of recruits in World War I (and to a lesser degree in World War II) demonstrated that an alarmingly large fraction of seemingly robust American youth were unfit and unhealthy. These findings were well publicized and had a big impact at the time in business circles. Many employers, thinking that an assessment of the total health of their workers was a valid objective and that they should therefore be comprehensive, modeled their screening programs on general health screening recommendations or military physical examinations without much regard for the occupational relevance of the tests or the evidence. This perpetuated the routine collection of information that was intrusive on the part of the employer (such as pelvic examinations), uninformative (routine chest films, which continued in periodic health monitoring long after they were dropped for hospital admission), irrelevant to working conditions (such as history of sexually transmitted diseases), and potentially harmful (such as routine lumbar spine films, which continued into the 1980s despite unacceptable radiation exposure and lack of usefulness). Toward the end of the twentieth century, there was a general turning away from screening as a primary means of health protection in most developed countries, although legislated surveillance continues to be an important part of occupational medicine practice in some countries, especially France.

The information collected by these programs has been subject to abuse. Violations of confidentiality were common in the past. Managers often assumed that information collected on behalf of the employer was open for their inspection. Inappropriate medical screening protocols and lapses in confidentiality continue today, but for the most part, screening has become more rational and respectful of worker privacy. This evolution has occurred as a result of union pressure, stronger ethical guidelines, changing management attitudes, and stronger protection for health records. Even so, health information collected by employers is weakly protected under the law—because it does not automatically

fall under HIPAA—and is hardly protected at all in workers' compensation. This is a major source of distrust between workers and occupational physicians practicing on behalf of employers.

## **MONITORING AND SURVEILLANCE FOR INDIVIDUAL RISK**

Following the health experience of workers is termed, in general, "health monitoring." When the emphasis is on a particular disease or set of disorders for which workers may be at risk, the monitoring program is called "health surveillance." Terminology is important in this area in order to keep concepts clear, but it is often confused in practice. This chapter will use consistent terminology adapted for occupational medicine. The term "monitoring" may seem to be overused in occupational health but conveys the sense of continuous or periodic examination to detect trends and unanticipated events. "Environmental monitoring" is the periodic testing of the workplace environment for hazardous or excessive exposure conditions (see Chapter 4) and is complementary to "medical monitoring," which is a general term for following the health experience of a worker through medical examination and testing. Screening is, properly, the practice of conducting tests or examinations in order to search for disease that is not yet recognized, but in occupational medicine the term also has the implication of identifying characteristics of the worker that are related to fitness for duty (see Chapter 18) or early indicators of occupational disease that suggest protection has not been adequate. When surveillance programs are scheduled at intervals, screening at each interval is called a "round," and the program overall is called "periodic health surveillance."

Monitoring and surveillance of individuals are undertaken to detect disease early, when it is more likely to be treatable, which is known as "secondary prevention" (see Chapter 8). Monitoring and surveillance of groups is undertaken to assess and deal with risks that may ultimately be shared by others in the workforce, which is termed "primary prevention." A surveillance or monitoring program of any

type, whether at the level of the individual worker or the population, should do both and implies a commitment to do something in response to the findings.

The occupational physician is often called upon to provide occupational health surveillance for workers employed in high-risk occupations, in which there is the potential for exposure to lead, noise, asbestos, pesticides, known chemical carcinogens, or other chemical or physical hazards. Such programs may also be conducted for groups of workers with unknown and unpredictable levels of exposure, such as employees in a given part of a plant where hazardous materials are used or chemical waste clean-up crews that are assigned where they are needed and so may encounter many potential exposures.

## **Objectives**

The general objectives of monitoring programs are the following:

- To detect changes before clinical disease becomes apparent, in order to remove workers from further exposure
- To detect overt disorders early enough to treat them with an improved likelihood of complete recovery
- To identify new or previously unsuspected health problems
- To confirm the absence of health problems as a check to ensure the efficacy of control measures

Without effective surveillance, it might be easy to overlook a breakdown in the measures used to control exposure to a hazard or the inadequacy of exposure standards previously considered acceptable. Monitoring programs that result in no evidence of work-related disease are therefore just as important as those that do, because they confirm that measures to control exposure to hazards are working effectively. This is particularly important in the case of occupational diseases, which tend to be rarer than injuries and more difficult to identify.

One might think that occupational health problems could be effectively identified by the worker's own physicians during routine care, without systematic surveillance and monitoring procedures. In the real world, that seldom works. Occupational diseases are encountered only occasionally in a typical medical practice and are seldom immediately recognized for what they are. A worker's personal physician rarely sees a complete picture of the health experience of employees in a particular industry or plant. On the other hand, a structured medical monitoring program that evaluates workers systematically can identify problems.

Health monitoring can also identify unexpected problems due to inadequate control of exposure. Exposure controls are not always universally protective. Some workers in some situations may be affected despite compliance with the standard, or the standard may be inadvertently exceeded. Monitoring compliance with exposure standards suggests how the worker ought to be doing. Medical monitoring informs occupational health professionals of how workers exposed to a hazard are actually doing. The alternative to "medical monitoring" is therefore to trust blindly in technical control measures and the opinions of experts, not knowing if standards turn out to be inadequate or inadvertently exceeded.

When the risk of a particular disease outcome is known to be elevated in a particular industry, a strategy to determine the group experience with that outcome is properly called "surveillance." When a specific health outcome is known or suspected, the task of following a population's health experience is greatly simplified. Surveillance programs are targeted to specific high-risk groups as defined by workplace assignment, known exposure history, and exposure-measurement data.

When a particular outcome is not necessarily known and the overall health experience of the group is to be observed, the strategy is more general and is properly called "monitoring," a more general term. In practice, the terminology is often mixed, but it is useful for this discussion to keep the concepts clear. "Medical monitoring," a term that has largely fallen into disuse, referred to using medical

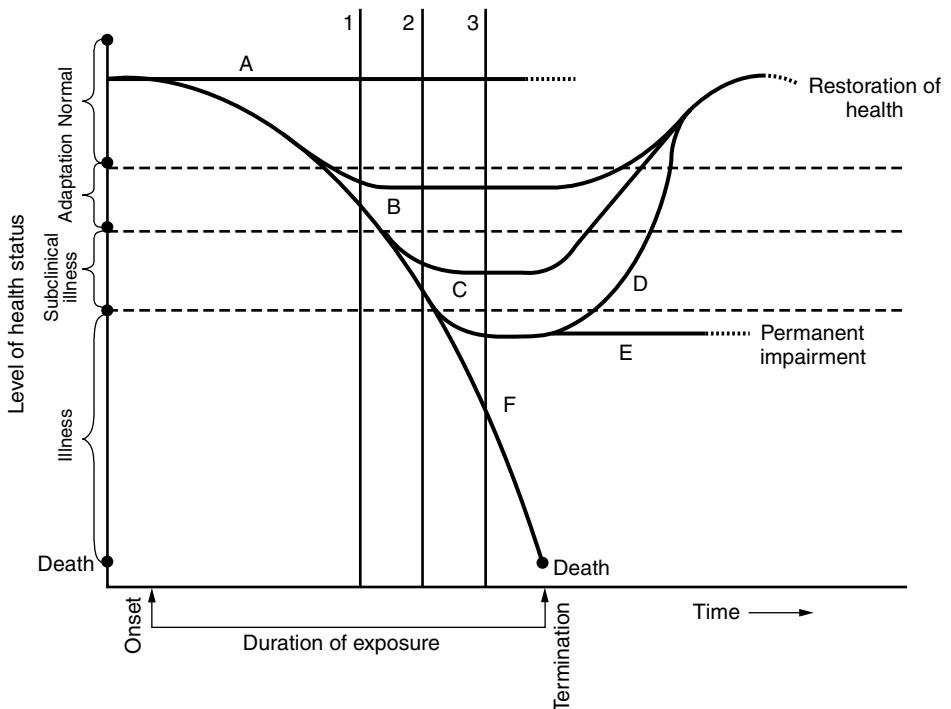
examinations and clinical tests to monitor the health of workers. However, the standard physical examination and routine clinical tests are insensitive for this purpose, as discussed in Chapter 1.

Surveillance is useful when the conditions in question are common or specific to a particular industry, such as the asbestos-associated lung diseases. Surveillance programs are usually restricted to high-risk groups and, when required by law or occupational health standards, called "mandated surveillance," are provided at the employer's expense to workers exposed to hazards such as asbestos, noise, and lead. Surveillance depends on the accurate identification of new cases of the disorder of concern among workers at risk. Monitoring depends on complete ascertainment of the health outcomes experienced by a defined population. Monitoring lends itself more readily to the discovery of previously unrecognized health problems. However, until recently the most common form of health monitoring was the annual physical examination, required by some employers. Although a reasonable tool for health promotion (see Chapter 19), the annual physical examination alone is inadequate for occupational health monitoring and so has become largely discredited as a health-monitoring strategy. Surveillance programs do not require as much collection of health data on the population as monitoring programs, and they are usually much less expensive.

## Principles

Figure 5.1 demonstrates the possible outcomes that follow an exposure, which may be chemical, physical, or psychosocial, but usually not biological. (Infection and immune responses function somewhat differently. Tuberculosis does fit the model, in part.) After an exposure is sustained, there may be no effect (curve A). Another possibility is that there may be an adaptive effect in which the homeostatic mechanisms of the body compensate for the effect of exposure but are not stressed beyond their capacity. For example, low-level exposure to certain aromatic hydrocarbons may induce metabolizing enzymes (curve B). A reversible subclinical effect (curve C) returns to normal when the exposure ceases. When the

exposure is sufficiently high, sufficiently long in duration, or cumulative, clinical illness may result, leading to a temporary illness (curve D), permanent disability (curve E), or death (curve F). One goal of any surveillance or monitoring program is to detect an effect in the adaptive phase (line 1), or at least in the subclinical phase (line 2 at curve C), before clinical illness develops. Another



**Figure 5.1.** Occupational health monitoring and surveillance and the course of a disorder (idealized). A represents baseline good health; B represents the response to an exposure (which may be a chemical, microbial, or other health hazard) with a biological effect resulting in adaptation; C represents a subclinical effect of exposure that does not result in toxicity; D represents disease and recovery; E represents disease and permanent impairment; and F represents disease progression. Individual health surveillance at time 1 might miss the onset of disease; at time 2, early disease might be detected; at time 3, disease is detected too late to prevent permanent impairment or progression. If the disease depends on continued exposure, as might be the case for toxicity, early recognition and removal from exposure may allow recovery. Adapted from "Occupational Health Monitoring and Surveillance," February 1985, *American Family Physician*. Copyright © 1985 American Academy of Family Physicians. All Rights Reserved.

goal is to detect illness at an early stage, when it may be reversible with prompt treatment. In Figure 5.1, this is represented by line 3 at curve D. Once the effect is recognized, exposure may be stopped and normal function regained, as in curve D. If the system fails to identify a hazardous condition early enough to prevent clinical illness, one at least hopes to limit disability (curve E). Progression to overt disability or death (curve F) is obviously unacceptable and represents a failure of the system.

A screening test must detect as early as possible an abnormality related to exposure if it is to be useful in disease prevention. Ideally, such a test would detect an adaptation to the exposure well before subclinical impairment develops, that is, as close to line 1 in Figure 5.1 as possible. Such tests fall into two categories: biological (effect) monitoring and toxicological screening. Biological monitoring, strictly speaking, involves techniques to determine the magnitude of an effect the exposure is having on the body without measuring the toxic substance directly. Measuring serum and red cell cholinesterase levels after low-level exposure to organophosphate insecticides or free erythrocytic protoporphyrin levels after exposure to lead are two examples of biological monitoring (see Chapter 4). Toxicological screening, on the other hand, includes tests to determine the levels of a toxic substance or its residues in tissue, body fluids, or excreta by direct measurement. Blood lead levels, urinary styrene, or benzene excretion are common examples of this more traditional approach. In both cases, the intent is to detect potentially toxic exposures before they manifest their toxic effects.

Often, useful early detection of the effects of a hazardous exposure is not possible during the subclinical phase. The outcome may be unchanged despite early recognition through screening. Some diseases, because of their natural history and biological characteristics, are too far advanced at the time detection is possible for a targeted intervention to change the outcome in all but a handful of cases. Unfortunately, this is the situation for sputum cytology and the conventional chest film when used in screening for lung cancer,

which is why neither modality has been adopted for surveillance purposes.

### **Test Performance**

The selection of a screening test is based on three variables: the “sensitivity” of the test, its “specificity,” and the “prevalence” of the disease in the community. These terms describe essential concepts that also apply to the use of any diagnostic tests in the recognition of disease. Sensitivity is the proportion of truly diseased persons in the population who are identified as diseased by the test. Specificity describes the intrinsic capacity of the test to separate out true cases of disease from false cases, which may, for technical or biological reasons, result in a positive test in the absence of disease; it is expressed as the proportion of truly nondiseased individuals in the population who will test negative. Prevalence refers to the frequency of the disease in the population at the time of the test. If the disease is very uncommon in the population at any given time, it is called “rare.” Table 5.1 shows the relationship among these variables.

The higher the sensitivity of the test, the more completely diseased individuals in the population will be identified. The higher the specificity of a test, the more reliably it will exclude nondiseased individuals. A test with a low sensitivity but a high specificity will detect only a small fraction of diseased individuals, but a positive test will be a reliable indication that disease is present in any individual; on the other hand, there may be many false negatives. A negative test result in this situation will not reliably rule out the disease. A test with a high sensitivity and low specificity will correctly identify most true cases, but it will also test positively for many individuals who do not, in fact, have the disease. A sensitive but nonspecific test yields many false positives, such that if a disease is rare in the population, the false positives may outnumber the true positives, requiring additional diagnostic tests to confirm the result. Tests that are both highly sensitive and highly specific are obviously best, and some tests, such as those based on radioimmunoassays, approach this standard. Most

**Table 5.1.** Measurements of Test Performance

Test Outcome	Disease Present*	Disease Absent*	Total
Test positive	a	b	a + b
Test negative	c	d	c + d
Total	a + c	b + d	a + b + c + d

a = Diseased individuals detected by the test (true positives)

b = Nondiseased individuals detected positive by the test (false positives)

c = Diseased individuals not detectable by the test (false negatives)

d = Nondiseased individuals detected negative by the test (true negatives)

\* In practice, presence or absence of disease is determined by comparison to a definitive, reliable test, which is often colloquially called the “gold standard.”

---

#### Performance Measures Independent of Disease Prevalence (Prev)

---

$$\text{Sensitivity (Sen)} = a/(a + c)$$

$$\text{Specificity (Spec)} = d/(b + d)$$

$$\text{False positive rate } (\alpha) = 1 - (\text{Specificity})$$

$$\text{False negative rate } (\beta) = 1 - (\text{Sensitivity})$$


---

#### Performance Measures That Depend on Disease Prevalence (Prev)

---

$$\begin{aligned} \text{Positive predictive value (PPV)} &= a/(a + b) \\ &= (\text{Sen})(\text{Prev})/[(\text{Sen})(\text{Prev}) \\ &\quad + (1 - \text{Spec})(1 - \text{Prev})] \end{aligned}$$

$$\begin{aligned} \text{Negative predictive value (NPV)} &= c/(c + d) \\ &= (\text{Spec})(1 - \text{Prev})/[(\text{Spec})(1 - \text{Prev}) \\ &\quad + (1 - \text{Sens})(\text{Prev})] \end{aligned}$$


---

tests in clinical use fall short of the ideal and thus require confirmation by additional studies. Tests that are neither sensitive nor specific are obviously useless but often attain a life of their own through custom, such as lactate dehydrogenase (LDH) assays, which were included in routine blood test panels in the 1970s, with no real benefit.

Together, the sensitivity and specificity of the test, in the context of the prevalence of the disease in the population, determine the

“predictive value” of the test, which is a measure of how useful it really is when applied to a particular population or group of workers. Predictive value is formally defined (by the *Dictionary of Epidemiology*) as the probability that a person with a positive result on the test will actually have the disease, and the predictive value of a negative result is the probability that a person in the negative result on the test will be free of the disease. For example, a routine chest film taken today in a healthy person from a low-risk population only rarely uncovers any unsuspected abnormalities and more rarely still identifies treatable disease. The test has a low positive predictive value in a community where there is little tuberculosis, and its high negative predictive value is not useful. A positive tuberculin test in a healthy person not immunized with BCG has a high positive predictive value for latent TB (but not necessarily active disease) and a high negative predictive value, unless the subject has a reason for immune impairment. When a tuberculin test yields a positive result, the worker being screened has been identified as having been exposed to a mycobacterium, the most likely of which is the tubercle bacillus, and therefore has a much higher probability of active tuberculosis than workers who test negative. Chest films performed on workers or patients who test positive in a tuberculin test are therefore much more likely to have active TB, and the positive predictive value for tuberculosis of a chest film in that situation will be much higher as a confirmatory test than when the chest film is used as a screening tool. These are familiar concepts to most clinicians.

Many common clinical tests—such as the stool guaiac (for blood in stool, indicative of colon cancer), blood pressure determination, and (when performed early as part of a program of serial measurements) spirometry—that are used for screening are also used for early detection of hidden disease. Although the technology is the same, there is a great deal of difference in the value of these tests when they are applied for screening rather than diagnosis. Most conventional clinical tests as used in medical practice are relatively crude when used for the purpose of detecting early signs of disease or exposure. That is because they were not designed for this purpose. They were designed to make

a diagnosis in sick people, who already have a high probability of having the disease. In other words, they are designed for specificity, not sensitivity. The diagnostic yield is naturally high in patient care because these tests are ordered for a small population of persons with a high prevalence of the disease, selected because they have symptoms or a history suggestive of disease. When such tests are used for surveillance or monitoring purposes, the diagnostic yield is low because almost all of those tested show no evidence of disease (for the most part, they would not be in the workforce if they were symptomatic) and so are predictably normal. The low prevalence of disorders in the population screened makes the positive predictive value low. The cost of finding a single abnormality indicative of an occupational disorder in one subject through screening a large population may be very high compared to the yield of the same tests in clinical situations in which sick people or people for whom there is a strong suspicion of disease are tested.

Surveillance programs should be built around the best available testing procedure in established use at the time they are put into place. A test that is difficult to interpret or that varies greatly from tests applied in the past will lead to confusion and uncertainty over long-term trends. A test that is novel or innovative may not be easily interpreted and may turn out to be unreliable. For example, in the early twentieth century there were several relatively elaborate tests for benzene exposure, none of which were superior to the white cell or platelet count, which worked only because exposure to benzene at toxic levels was very common. The most widely applied mandated surveillance procedures are based on relatively simple and straightforward tests: the lead standard on blood lead, the noise standard on audiometric screening, and the asbestos standard primarily on the chest film. A surveillance program may incorporate an experimental test, but any test of uncertain validity must be in addition to recognized standard tests for the disorders being sought and given with the consent of those being tested. Likewise, when a new test is validated and introduced as a replacement for an older test of less reliability, both tests should be performed together for several years so that the results may be compared and trends will not be obscured by the transition.

For reasons of cost, sustainability of the screening program, and reliability, occupational health and surveillance programs are designed to be as simple as possible and to rely on a few well-established tests. The surveillance program for noise-induced hearing loss for workers exposed to high noise levels, for example, relies on an annual audiogram, covering prescribed frequencies and performed by certified personnel in a very standardized manner on equipment with prescribed specifications. Audiometry is a screening test that generally works well for its intended purpose. An example of a screening test that does not work so well is spirometry, when it is used to evaluate workers exposed to silica or asbestos. The test simply does not work as an early indicator of silicosis or asbestosis, because restrictive changes do not occur until the process is advanced. Even so, it is written into OSHA regulations and so is a required part of mandated surveillance for silica- and asbestos-exposed workers. It does have the value of quantifying lung function for later reference in following the worker-patient.

Table 5.2 summarizes the essential characteristics of any test that is considered for a surveillance or monitoring program. Because screening tests are applied to people who are not likely to have a serious illness, they must be safer and more acceptable than clinical tests, which are done for a different purpose.

**Table 5.2.** Criteria for the Selection of Screening Tests for Monitoring and Surveillance Programs

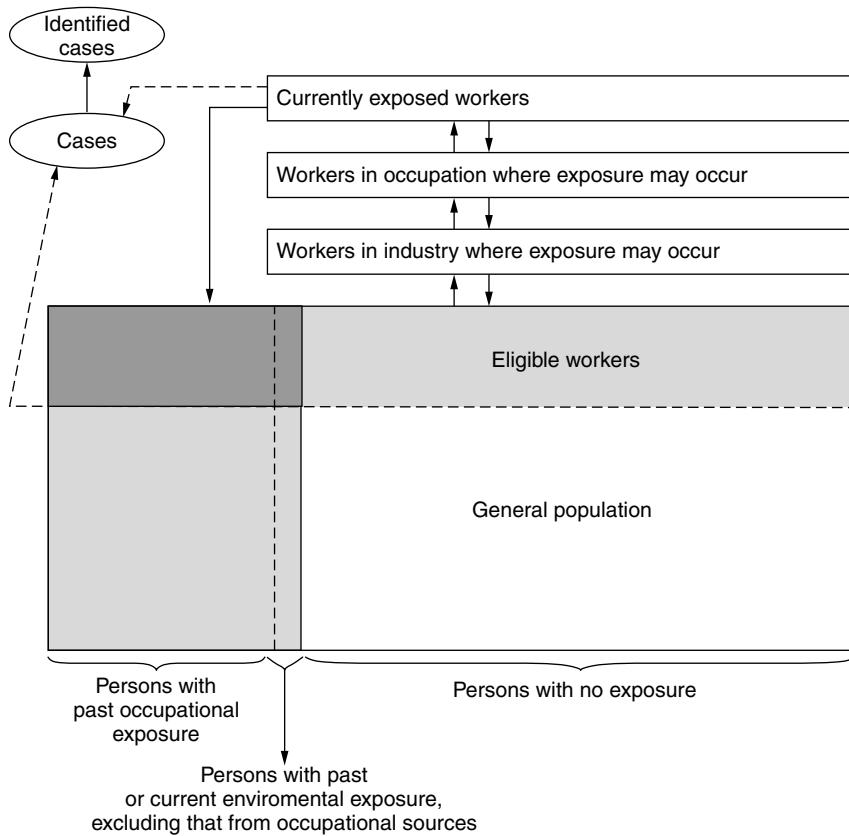
- 
1. The test should be sensitive and specific.
  2. The test must be simple and cheap.
  3. The test must be very safe, to avoid causing more harm than good by subjecting large numbers of normal people to a small but real risk.
  4. The test must be acceptable (i.e., not inconvenient, time-consuming, uncomfortable, or unpleasant) to the subjects and to those conducting the test.
  5. The test should be capable of detecting disease at a time when intervention is likely to result in cure.
-

The frequency of testing required in a surveillance program depends in part on the biology of the disease. When there is no obvious periodicity to be preferred, most surveillance procedures are repeated annually, because one year is a convenient time period and is easily tied to the workers' birthdays, dates of hire, or the business cycle. The noise standard, for example, requires annual audiometric screening for workers exposed to 85 dB or more of noise. There is nothing inherent in noise-induced hearing loss that makes testing every year optimal, but testing less often may lead to undetected progression of hearing loss between tests and to tests being inadvertently missed. Testing more often would be expensive and unlikely to identify a much higher proportion of new cases. Many employers send their employees for screening in groups during the off-season in business; others conduct screening programs throughout the year and have employees screened on the anniversary of their employment or reassignment. The periodicity of other tests may be less logical. The present OSHA asbestos standard is intended primarily to detect asbestosis, the fibrotic lung disease associated with asbestos exposure, which takes many years to develop. The latency periods for asbestos-associated cancer (bronchogenic carcinoma or mesothelioma) are much too long and the cancers have too poor a prognosis to be effectively controlled or mitigated by early detection and treatment.

## **IMPLEMENTING MONITORING AND SURVEILLANCE PROGRAMS**

Monitoring and surveillance amount to a process of identifying and then confirming prevalent cases of disease within a population in which workers and other residents of the community share many determinants of health and even exposures that may be associated with the disease outcome. One way of looking at surveillance in the context of occupational disease is illustrated on a population basis in Figure 5.2. The worker is a member of a workforce with more or less common exposures and more or less common health risks. The

workforce, in turn, is drawn from the community, which has its own prevalence of morbidity and disease patterns. However, most workers are drawn from among the healthier members of the community. The workforce is hired from a pool of eligible workers within the general



**Figure 5.2.** Schema for understanding the workforce and its relationship to the general population, with respect to work-related exposures. The workforce is drawn from the community and therefore reflects many of its health characteristics. However, the workforce is also highly selected and shows differences in health experiences, exposure, and health risk. Not every worker in a particular industry is exposed to the characteristic risks of that industry; it depends on the worker's specific occupation. Some members of the general population, usually a small number, may share exposures due to other jobs, passive or community exposures, and environmental exposures, depending on the hazard.

population, which already includes people with past occupational exposures, including some relevant to the work. Some of the workers are hired in an industry and in specific jobs or occupations where an exposure may occur. This is the pool of currently exposed workers. From this pool come the cases that arise from work. For many disorders, such as lung cancer or asthma, there will be other cases that arise from the general population or prior exposures in other jobs. Some of these cases remain unidentified. Other cases are identified and reported in surveillance and monitoring, and others become the patients in the healthcare system.

A system of monitoring or surveillance for occupational disease is a backstop for measures to control the hazard, not a replacement for those measures. A medical monitoring or surveillance program is also a commitment to action. Abnormal findings must be acted upon, not filed away and forgotten. A medical monitoring program is not a substitute for action to control workplace exposure. It is a commitment to being sure that the worker is protected, in fact as well as in theory. Protecting the health of workers by controlling exposure is the primary objective, and medical monitoring is the next essential step to protecting the worker's right to a safe workplace.

Even given the relatively crude clinical tests used in most screening protocols, occupational physicians still may identify occupational disorders in workplaces and situations in which exposure was thought to be controlled. In a workplace thought to have had adequate controls in place, it is not unusual for workers to be discovered as having unsuspected abnormalities, demonstrating that existing controls on exposure are not working. Without medical monitoring programs to confirm the effectiveness of control measures, workers continue to be at risk, and occupational disorders go unrecognized, uncorrected, and uncompensated.

It is worthwhile to generate a flow sheet on the findings in each employee's individual record and to provide a copy of the surveillance protocol or the standard, if mandated, if the case is referred. Periodically, data for all high-risk employees under surveillance should be reviewed together by work unit to identify trends and

**Table 5.3.** Population-Monitoring Studies of Historical Importance in Occupational Epidemiology

<b>Study</b>	<b>Data Type</b>	<b>Population</b>	<b>Era</b>	<b>Risk</b>
<i>Mortality</i>				
Guralnick [1963]	Census, death certificates	U.S., males	1950	SMR
British Registrar-General [1961,1971]	Decennial census, death certificates	England and Wales (only), total	Series, 1971 most recent	SMR
Milham [1976]	Death certificates	Washington (state)	1950–71	PMR
Petersen and Milham [1980]	Death certificates	California (state)	1956–61	PMR
Dubrow and Wegman [1984]	Death certificates	Massachusetts (state)	1971–73	OR
Rhode Island [1981]	Death certificates	Rhode Island	1968–72	SMR, PMR
Copplestone [1967]	Death certificates, census	New Zealand, males	1959–63	SMR
<i>Morbidity</i>				
British Registrar-General [1972,1975]	Cancer incidence	England and Wales (only), total	1966–67, 1968–70	PMR
Roswell Park [1977]	Hospital registry	Referred patients (western New York state)	1956–65	OR
Social Security [1980]	Disability claims	U.S., males	1959–62	PMR
Society of Actuaries [1967]	Insurance claims	U.S., employed	1967	NA
Third National Cancer Study [1971]	Cancer incidence	U.S., total	1970	OR
Health Interview Panel	Survey	U.S., total	1969–74	NA

SMR = standardized mortality ratio

PMR = proportionate mortality (or morbidity) ratio

OR = odds ratio

NA = study not available for review

patterns that might otherwise be overlooked and to identify workers who may show suggestive or borderline results—such as rising but not overtly elevated blood lead—or abnormal loss of capacity—such as workers who lose lung function or hearing at a rate greater than expected, even if they are within the normal limits for the test.

If possible, environmental monitoring data from industrial hygiene surveys should be made available to the clinician and must be disclosed to the employee. In practice, environmental monitoring data is seldom provided to outside physicians, except when brought along by the worker-patient him- or herself. This is not good practice.

A major challenge facing most physicians in conventional monitoring programs based on clinical evaluation is maintaining a high level of attention and alertness to clinical signs while examining numerous individuals, most of whom are healthy. Physicians often find it difficult to maintain their interest and index of suspicion in the face of repetitive negative examinations. The number scheduled during any one visit should be limited, if possible. As many of the examination components as possible should be delegated to professionals with an interest in process as well as medical management. Occupational health nurses, for example, often do better on repeated examinations than physicians because their interest and motivation remains high. Their training emphasizes the importance of the process of occupational health monitoring, whereas that of the physician emphasizes the detection of abnormalities. Physicians usually find the absence of abnormalities boring and risk becoming distracted or going through the motions automatically; even excellent clinicians may miss low-grade positive findings when they are encountered without warning after a long line of normal subjects.

Surveillance programs may be conducted in-plant or by an outside medical facility or practitioner under agreement with the employer. In establishing such an agreement, the tests to be conducted, quality assurance standards, and the frequency of screening must be clearly specified, and procedures for managing the cases of individuals found to have a disorder must be formulated in advance. Unfortunately, most primary care providers do not take periodic health surveillance programs very

seriously and often conduct very cursory examinations at best. They often do not read and do not comply with instructions. One reason for this is that in the past, many periodic health surveillance programs were poorly designed and were often just routine annual medical examinations, with obviously irrelevant tests (such as pelvic examinations) and very low yield. For this reason, periodic health surveillance should normally not be referred to the worker-patient's own primary care provider, and if performed under a contractual arrangement, it should specify quality controls, external review, and accountability.

Even when mandated by law or contract, a surveillance program should have the approval and understanding of the employees involved, achieved through their union or through direct worker representation. There should be an agreement that explicitly states the purpose of the program, the measures being taken to control workplace exposure, measures each employee can take to minimize personal exposure or to reduce the risk of the outcome in question, the right of each employee to withdraw from a voluntary program, and the contractual obligation of the worker to participate in mandatory programs and those required by the employer as a condition of employment, and that specifies an authoritative source (usually the supervising physician), who will be made available to answer inquiries. Employee attitudes and satisfaction with the surveillance program should be determined regularly so that problems can be resolved. Unexpressed dissatisfaction usually reveals itself as a refusal to participate in a voluntary program or a passive failure to comply with a mandatory program. Employee doubts about a screening program usually hinge on concern over the confidentiality of medical records. They may also reflect the suspicion that drug testing will be performed at the same time.

## **ETHICS AND SOCIAL ISSUES**

Although less of an issue today, in the 1980s there was a general questioning of the rationale for monitoring and surveillance programs, particularly in Canada and the United Kingdom. Some

suggested that there is a potential conflict between individual human rights and the conduct of monitoring, surveillance, and other medical evaluation programs that gather information on employees. The thrust of this argument was that abuses had occurred and would likely occur in the future, and that guidelines for the conduct of monitoring were needed. Most of these issues related to the confidentiality of medical information and the risk of discriminating against workers with abnormal results. This debate led some parties to propose stringent guidelines for the regulation of occupational health monitoring and surveillance activities. This concern arose out of an attitude of profound skepticism over the efficacy of surveillance procedures and a concern for the possible coercion of workers into participating. A consensus emerged on several points among occupational health professionals following this issue:

- The conduct of outcome monitoring and surveillance programs for working populations of interest is not an acceptable substitute for exposure monitoring and control.
- Employers, health professionals, and any other parties involved in monitoring or other health evaluation programs have a moral and ethical obligation, and often a legal duty, to see that the workers under evaluation understand the purpose, degree of validity, and consequences of participation.
- Workers have an absolute right to be informed of their individual results and the implications for their health and well-being; this information is confidential.
- Employers have no right to information that does not pertain directly to either occupation-related disorders or the capability of a worker to perform the job assigned.
- Unusual or special evaluation techniques—such as a biomedical monitoring procedure under development and not yet standardized—especially if invasive, should be conducted only with the specific consent of the worker.

- Commitment to a monitoring or surveillance program implies that the data will be gathered for a purpose and that the findings will be acted upon.
- Data from medical monitoring and surveillance activities should not be used as grounds for personnel actions, such as laying off workers.
- Employers should be responsible for the cost of the program because they are ultimately responsible for protecting the health of their workers.

Although these eight points represent substantial areas of agreement among occupational health professionals, labor representatives, and employers, many other issues remained in dispute.

However, events superseded these arguments. Medical activities in the workplace were undergoing profound change at the time, and much of this change was due to new information generated by research and monitoring activities and new ways to use the information, such as risk assessment. Clinical epidemiology and occupational medicine research provided insights to refine medical monitoring procedures and to make them more objective. Criteria for assessing the validity and usefulness of tests emerged from rigorous epidemiologic research. Unfortunately, many regulations and programs followed in industry are based on the technology of the 1950s, frozen in time while medicine has moved forward.

Today, a combination of improved technology, greater protection for medical records, case law better defining the right of employers to enforce occupational health protection, and greater familiarity with monitoring procedures have made surveillance and monitoring programs less controversial. Objections are still often raised, especially when there is confusion over periodic health surveillance and drug screening. At present, OSHA generally imposes a requirement on management to provide mandated surveillance, but does not impose a requirement on the worker to participate.

## POPULATION-BASED MONITORING

Monitoring and surveillance data at the population level are essential for public policy, to document trends, and for epidemiological purposes. Without such data, the magnitude of problems cannot be assessed, improvement or deterioration cannot be ascertained, and problems cannot be localized. However the collection of data is costly, requires a bureaucracy, and is not politically attractive. For these reasons it is often neglected, and sometimes intentionally so, because population-based monitoring may document inconvenient facts.

There are two general approaches to monitoring, active and passive. Active monitoring refers to active case-finding, such as requiring automatic reporting from laboratories for positive tests. Passive monitoring relies on the voluntary notification by physicians and other healthcare providers of new and suspicious cases. In practice, passive monitoring does not work well except for acute diseases of indisputable public health significance and where reporting is required by law.

Population health surveillance is an integral function of public health. Most such programs rely on existing information sources, such as population-based cancer registries and (in Canada) healthcare utilization data, supplemented by occasional special surveys or local studies. Large-scale surveys for the collection of new health data on populations are also undertaken from time to time, such as the National Health and Nutrition Examination Survey and the recently launched National Children's Study in the United States. There is currently a great deal of interest in "tracking" exposure to chemical hazards over time in the general population (see Chapter 2). Such studies are difficult to organize, are expensive, and usually require new infrastructure.

### Surveillance for Occupational Health Trends

In the United States, information on occupational disease and injury is collected from a random sample of employers and then compiled by the Bureau of Labor Statistics (BLS), which publishes an annual report. This source is valuable in documenting trends but is aggregated at too high a level for most purposes. Because it relies on

reported or recognized cases, the BLS data considerably underestimate the scope of occupational diseases. BLS also publishes an annual Census of Fatal Occupational Injuries.

State agencies have mandatory requirements for reporting occupational injury and disease, but compliance is not enforced. Workers' compensation data may also be tracked through state insurance commissions or state or provincial workers' compensation boards. Data from these sources show a substantial undercount for occupational disease and chronic conditions.

Population health surveillance can also be undertaken using the existing medical, public health, and occupational health infrastructure. See Chapter 3 for a discussion of historical population-based monitoring programs for occupational injury and disease, based on publicly registered death certificate data and disability data from social insurance programs. (See also Table 5.3.)

These studies have a venerable history, especially in the United Kingdom, where the British Registrar-General's Decennial Reports were once the standard resource for assessing risk of mortality by cause associated with occupation. In North America, a number of studies of shorter duration were conducted, mostly at the state level, pioneered in the 1970s and 1980s by Samuel Milham in the state of Washington. Earlier efforts were directed toward the U.S. population as a whole, including a massive three-way analysis by Lillian Guralnick—which was unfortunately limited by restricted coverage of health outcomes—and a study on disability using Social Security data.

These surveillance data resources were immensely useful. They suggested hypotheses for testing using specific studies, and in the absence of a targeted study they provided at least partial guidance for the assessment of health risks. They had other highly desirable features, among them:

- They defined a general population that was comparable to the working population.
- They presented the data on risk in a way that could be interpreted as an elevation above the level of the general population.

- They were longitudinal in design and could, in theory, demonstrate trends over time or adjust risk estimates as they changed.
- They made data accessible to any user through technical reports.

These studies were casualties of decreasing budgets and declining interest. They were subject to all the usual biases of misclassification, made worse by inaccurate reporting of occupation on the death certificates among the mortality studies. In general, these biases tended to underestimate risks. The problem of multiple comparisons also made individual findings suspect unless they were consistent with other such studies or confirmed by targeted investigations. The emphasis on mortality made interpretation difficult for disorders that were disabling but seldom fatal. Even so, the advantages of these surveillance studies were considerable. Mortality by cause of death was by far the most common endpoint because, for all the uncertainties in death certification, it was a more reliable description than morbidity and was a useful surrogate for incidence or morbidity for serious illness.

What is needed today is a database, much like that of the occupational surveillance studies, on a larger population base. Gilbert Beebe made this recommendation in 1983, but no progress ensued. It should be continuously renewed, preferably through some system of ongoing surveillance, so that longitudinal trends can be tracked. Ideally, it would have the following characteristics:

- Comprehensive coverage of a defined population, with the capacity to capture individual demographic characteristics
- Reasonably accurate determination of diagnosis to support reasonably certain estimates of incidence, morbidity, and mortality
- Accurate reporting of both occupation and industry
- Accessibility to users, such as workers' compensation agencies
- Audit and quality assurance

- Capability of linkage to other data sets, such as cancer registries and healthcare utilization data
- Strict confidentiality with respect to individual identifiers
- Compatibility with extant coding systems and conventions, such as ICD-10
- Continuous reporting for ongoing surveillance, with automatic data updating and reporting
- Interpretive features to ensure valid use of the data

Additional features that would be desirable but probably not practical include some surrogate measure of individual exposure level (such as duration of employment), impairment level for living subjects (to support estimates of disability), and multiple occupations and avocational activities (such as hobbies and volunteer work). Capturing individual exposure level would not be necessary, because a job exposure matrix based on hygiene data could be built into the system and would provide a valid estimate for a population. Obviously, the practicality of adding these features is limited by the degree of intrusiveness and cost.

Data linkage on this magnitude is technically possible but beset by practical problems and issues of public policy. Such a comprehensive system may not be possible today, particularly under privacy legislation in many countries in Europe. However, it should be possible to develop such a system in a few jurisdictions to serve as models and as data resources. Indeed, data linkages such as those described are already under discussion in Canada to support administration and resource utilization among the provincial health insurance schemes. Workers' compensation agencies have not been a major part of these discussions despite their obvious relevance.

Given the large issues and substantial obstacles that would block early development of such a system, one may propose an interim arrangement to at least improve the quality of information currently used as the basis for adjudication and social decision-making involving injured workers:

- Data sharing among workers' compensation and insurance carriers on an ongoing basis through a central statistical clearinghouse, perhaps required by state insurance commissions (highly problematical because claims data are proprietary)
- Actuarial analysis of pooled claims data on a much larger scale, beyond the scope of any one insurance carrier
- Analysis of data, taking into account employers' performance and socioeconomic status (because the business of insurance carriers is often segmented to different levels of the market, reflecting characteristics important in assessing risk)
- Changes in legislation to declare data a public resource to be accessed in the public interest (which is unlikely even to be considered in the absence of a national health insurance system and pressure for quality improvement in general healthcare within that system)
- Protection for competitiveness among insurance carriers, with universal access to the pooled data (forcing competitiveness on the basis of efficiency rather than proprietary data)
- Access to these data by legitimate investigators from academia, industry, and labor (but not to data specific to one company in a scheme where the private sector participates)
- Support by industrial insurance schemes, perhaps through payment into a pooled fund, for an ongoing program of surveillance, cross-jurisdictional (and international) comparisons, and selected targeted studies
- Use of these data in setting priorities among occupational health regulatory and research agencies (notwithstanding that other considerations also affect priorities)

At the very least, workers' compensation and disability insurance data can be used more effectively in the assessment of occupational risks. For example, David Goldsmith has demonstrated that workers' compensation data in California on silica-exposed workers yielded estimates of

risk for lung cancer similar to those of more conventional cohort studies. These data are not easy to work with and have problems of bias; however, these problems should serve as a challenge to epidemiologists rather than a reason for neglecting this rich source of data.

### **Sentinel Events**

Another approach to monitoring the occupational health experience of a community is through the identification of “sentinel events.” Sentinel events are cases that strongly suggest the presence of a larger occupational health problem in the community. For example, identification of a case of pleural mesothelioma strongly suggests the presence of other asbestos-related diseases in the population. Because most occupational diseases are grossly underreported, the identification of clear cases usually suggests a bigger problem in the community.

Sentinel-event notification provides a unique opportunity to identify cases, situations, and trends indicative of a changing pattern in the community. One approach to developing this capability is the Sentinel Event Notification System for Occupational Risks (SENSOR), operated by the U.S. National Institute for Occupational Safety and Health. This system is based on notification of sentinel events and collection of the data. A major advantage is that reporting does not have to be complete or cover the entire population to yield useful information.

The SENSOR network is a system for routinely identifying and reporting designated “sentinel events (occupational)” (the sentinel event concept can also be applied in nonoccupational disorders) that consist of cases that are (1) exclusively or predominantly occupational in origin; (2) known to reflect the presence of significant uncontrolled exposure in the population at risk; (3) unequivocal in diagnosis; and (4) known to be associated with a proportional fraction of subclinical, undiagnosed, or unrecognized disorders likely to come to medical attention. SENSOR depends on reporting from the individual states, however, and does not completely cover the

United States. Initial efforts were focused on six sentinel events: silicosis, occupational asthma, pesticide poisoning, lead poisoning, carpal tunnel syndrome, and noise-induced hearing loss. Preliminary results from the demonstration project for occupational asthma suggest that the system can be effective in identifying new outbreaks, changing trends, and creating opportunities for early intervention. Also, preparation for SENSOR has had additional benefits in forcing more consistent diagnostic practices among the participating providers.

Ideally, it should be possible to estimate the prevalence of "hidden cases" in the community from reports of sentinel events, if referral patterns are known and the proportion of cases arising can be determined. For example, the detection of a case of pleural mesothelioma can be assumed to indicate the probable presence in the population of a proportionate number of other, probably undetected cases of asbestos-related disease. Tracing the frequency and distribution of these sentinel events may substitute effectively, quickly, and cheaply for more-elaborate surveys. If its validity can be demonstrated, the technique promises to provide an effective, rapid, and inexpensive alternative to traditional methods. The system assumes that cases are generated by a base population served by a "sentinel provider," or healthcare institution encountering patients likely to have manifestations of exposure-related disorders. These sentinel providers may include (1) occupational health clinics, (2) poison-control centers, or (3) potentially, individual practitioners. Hospitals are generally unsatisfactory for the purpose, because few occupational disorders require admission for evaluation, except cancer. Each type of provider presents certain advantages and disadvantages.

The ideal provider for this purpose would be a health service covering a well-defined and representative population, to which all cases of a known or suspected occupational disorder would be referred by primary-care physicians sufficiently trained and knowledgeable to catch the vast majority of cases of interest. In actual practice, providers serve an ill-defined network of cooperating practitioners, each of

whom has a variable threshold, or “index of suspicion,” for presuming that a given disorder may be occupational or environmental.

Referral centers receive cases that are prescreened and may be inconsistent or biased in referral patterns. Knowledge of a sentinel provider’s referral patterns and catchment area is important for interpreting how well trends reflect what is happening in the base population.

## **RESOURCES**

- Ashford NA, Spadafor CJ, Hattis DB, Caldart CC. *Monitoring the Worker for Exposure and Disease: Scientific, Legal, and Ethical Considerations in the Use of Biomarkers*. Baltimore: Johns Hopkins Press; 1990.
- Mitchell FL. *Instant Medical Surveillance: The Evaluation of Biological and Chemical Dangers*. 2nd ed. Beverley Farms, MA: OEM Press; 2007.

## **READINGS**

- Beebe GW. Prospects for national data on occupational mortality. In: Chiaze L, Lundin FE, Watkins D. *Methods and Issues in Occupational and Environmental Epidemiology*. Ann Arbor, MI: Ann Arbor Science; 1983:75–82.
- Goldsmith DF. Use of workers’ compensation data in epidemiology research. *Occup Med: State of the Art Rev*. 1998;13(2):389–416.
- Guidotti TL, Baser M, Goldsmith JR. Comparing risk estimates from occupational disease monitoring data. *Public Health Rev*. 1987;15:1–27.
- Myers JE, Thompson ML. Meta-analysis and occupational epidemiology. *Occup Med*. 1998;48:99–101.

*This page intentionally left blank*

# **6 SAFETY AND INJURY PREVENTION**

Of all the disciplines in occupational health and safety, safety science has traditionally had the least to do with medicine, both technically and professionally. It is therefore overlooked in most textbooks and treatments of occupational and environmental medicine (OEM). However, many concepts in safety science have application in medical practice and occupational health, and the broader field of injury prevention is an important, but often unrecognized, part of environmental health management. There are also many lessons from safety science that could be applied to patient safety in healthcare institutions, although this topic will not be explored here. This chapter will not attempt to provide a systematic overview of safety science or community injury prevention, neither of which are usually responsibilities of physicians. Rather, it will describe key concepts basic to both and useful to the physician in understanding these closely related fields.

The terms “accident” and “accidental” are not used in modern injury prevention. “Accident” implies an element of chance; the preferred terminology among injury-control professionals is “incident,” in order to remove fatalistic implications. This can make for awkward terminology, and the term “accident” may be too entrenched in business

terminology to dislodge. Even so, this chapter will conform to the new convention of avoiding “accident.”

The elements of an incident include: (1) the person susceptible to the injury, (2) the hazard that is capable of inducing the injury, and (3) an environment that brings them both together in the workplace. Bringing together these elements creates the context for the incident. A person may decide whether or not to take a risk on the spur of the moment; a person may not perceive the hazard; a person’s appraisal of it may be faulty in some way; a person may be distracted; or a person may make a decision that in retrospect was not reasonable; but whatever occurs, there is a very critical period just before the incident when these things come together into an incident-facilitating situation. Most of those situations will result in a near miss, an injury that could have happened but did not occur. However, these elements come together with a certain probability: an event will happen, and the injury occurs. It is always unexpected going forward, but it always seems inevitable looking backward. It is always unintentional, and yet the circumstances are always to some degree predictable. This results in effects that may include: injuries, damage to property, financial loss, legal liability, and future costs and overhead.

Injury prevention is not entirely about preventing the event, although prevention is primary. Mitigation plays an important role. Once the injury occurs, a great deal can still be done to minimize its impact. Fair compensation follows in the management of the consequences of an injury that occurs in the workplace. This is discussed in Chapter 21.

Some basic principles of injury prevention are summarized in Table 6.1.

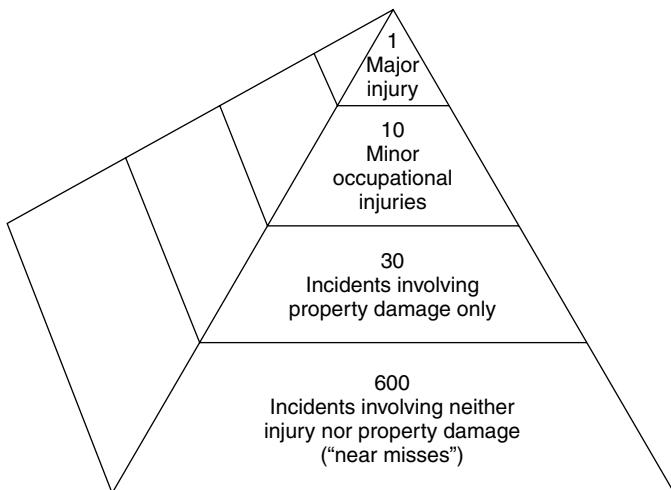
## THE INJURY PYRAMID

The “injury pyramid” is a graphic representation of the relative frequency of fatal, serious, minor, and near-miss accidents resulting in only property damage (see Figure 6.1). The injury pyramid is shaped differently for childhood injuries, motor vehicle collisions,

**Table 6.1.** Principles of Injury Prevention

- 
1. No one knowingly exposes oneself to an unacceptable risk.
  2. Adults must often take risks to achieve their objectives.
  3. Safety for an adult means a risk that is acceptable and that is minimized to the extent possible.
  4. Safety is not at odds with productivity.
  5. Every accident is the product of multiple factors.
  6. Injury prevention is risk management (see Chapter 7).
  7. An incident is a symptom of an underlying problem, usually something wrong with the management system.
  8. Certain situations invariably present a high risk of injury, but they can be anticipated and controlled: unusual or nonroutine work resulting in a departure from standard procedures, such as upset or emergency conditions, unusually slow periods, changes in the pace of work (speeding up or slowing down), and interruptions.
  9. Safety is protection against the abrupt and uncontrolled release of energy; it can be accomplished by: avoiding the presence of an energy source, separating the worker from the energy source, controlling the energy source, diverting or preventing the release of energy from the source, slowing down the release of energy from the source, and providing personal protection to the worker to prevent the energy release from causing injury.
  10. Safety should be achieved like any other business outcome: by the same means of planning, evaluation, and management.
  11. Every manager should be accountable to his or her supervisor for safety, and safety performance—including both process measures for injury prevention and outcome measures in injury experience—should be part of a manager's regular performance evaluation.
- 

and other types of injury. The occupational injury pyramid differs from company to company and from country to country and over time. However, for planning purposes, a robust rule of thumb is that for every fatality and seriously disabling injury, there are at least ten minor injuries resulting in some degree of inconvenience and minor impairment, but typically without permanent or even serious temporary impairment. For every one of these minor injuries,



**Figure 6.1.** The “injury pyramid” for occupational injuries and incidents. (Data for illustration only; actual ratios vary.)

there are at least three accidents in which there is only property damage in a workplace setting. There are about twenty near misses for every accident involving property damage or worse consequences. Thus, for every serious, disabling injury there are perhaps six hundred incidents in which there is no injury and no damage, but a near miss. Since this formulation was first put forward in 1969 by Frank E. Byrd, who worked for an insurance company, it has been confirmed with varying ratios in studies of occupational injuries across many countries. Thus the principle remains valid although the overall rate of occupational injuries is falling in most developed countries.

The injury pyramid shows quite clearly that liability, legal liability for serious injury, costs in property damage, costs in compensation, costs in medical care, loss in the sense of property damage, and risks that something bad will occur are all intertwined. If one prevents injury, one may also prevent legal liability and rising insurance premiums; one may also reduce workers' compensation assessments and minimize direct loss and property damage.

The pyramid of occupational injury also suggests ways to deal with the prevention of injury. One is to reduce the number of incidents so that the near misses do not happen, on the theory that fewer incidents will occur higher up the pyramid. Another approach is to prevent an incident in which property damage occurs from resulting in an injury by building in protection in the form of safeguards and personal preventive equipment, thereby limiting the damage resulting from an injury that does occur (and changing the shape of the pyramid at the base), providing the seriously injured with prompt attention, and minimizing the consequences of the event.

## **CONCEPTS OF SAFETY AND INJURY PREVENTION**

What can be done to change things? Injury prevention is one field in which practice follows theory. Safety practices on the shop floor are determined by the model adapted by management and workers of what safety is all about. It will or will not work depending on how well the nature of the problem is understood on theoretical grounds. M. D. Harvey (1984) has reviewed theories of injury (or, in older terminology, “accident”) causation in detail. His categories are summarized in Table 6.2. The following discussion takes a more historical or anthropological approach.

Society has gone through different ideas about injury causation and why injuries happen. These theories have profoundly affected social attitudes and have shaped thinking in safety science and injury prevention.

The first, relatively primitive idea of why “accidents” happen reflected a concept of fatalism. In societies around the world as well as in this society in previous times, a fatalistic point of view was the norm. “Accidents” were inevitable. Injury was a matter of being in the wrong place at the wrong time. A related notion was the traditional idea that accidents are “acts of God.” The problem with the concept of an act of God is that it tends to maintain the status quo. If no one can guess chance events beyond one’s control, no one feels a need to change very much. Society then moved on in its thinking,

**Table 6.2.** Some Historical Theories of Incident Causation and Their Present Status (Harvey 1984)

Theory	Contemporary Status
<i>Single-factor theories</i>	
The “accident-prone” worker (associated with Taylor)	Discredited, except insofar as a mismatch may exist between worker and task
Psychodynamic origin of incidents (not discussed in text)	Discredited, except in trivial sense in which conflicted behavior, carelessness, or distraction may contribute in a given situation
“Social-environmental model”: goals-freedom-alertness (not discussed in text)	Lacks credibility as complete explanation, but concept of worker autonomy and participation contributes to multifactorial models
“Human factors” (associated with Taylor)	Perceives risk of injury as a function of workplace and job design; also the basis of modern ergonomics
“Domino model”: factors sequentially conditioning the unsafe act	Simplistic, omits data; useful mainly as framework for an incident investigation
<i>Multiple-factor theories</i>	
“Public health” model: multifactorial model recognizing host-environment-agent	Useful in analysis; conceptual basis for modern models, including Haddon
<i>Systems approaches</i>	
“Human engineering”: perception-understanding response	Incorporated in Haddon’s Matrix and is the basis for contemporary theory
“Expectancy-skill-decision”: emphasis on information handling, reality checks, and information-receiving capacity	Useful in understanding incidents during deviations from routine or during slow pacing of work; used in ergonomics

but adopted an attitude that was in many ways worse, taking a step backward into a concept of “acts of man.” When this philosophy predominated, workers were unnecessarily blamed for their injuries, and sometimes, as with other misfortunes, injuries were viewed as punishment for sins. This notion took hold especially in injuries sustained at work, and attitudes that reflect it can still be encountered from time to time in the forms of “blaming the victim” and “self-blame,” neither of which is very constructive.

Liberal thinking in Europe in the 1700s, during the Enlightenment, and in the 1800s put emphasis on the rights and responsibilities of the individual person. As the Industrial Revolution progressed, the idea that later came to be called “social Darwinism” increasingly held sway. This concept, which predated Darwin’s concept of natural selection, held that if one were not fit, one did not survive, and that society was ultimately better off if the unfit perished. A *laissez-faire* attitude emerged, replacing the previous attitude of *noblesse oblige* toward the lower classes. Workers were expected to fend for themselves and to look after their own interests, including their own safety, although they lacked the means to do so. Women and children entered the labor market and worked long hours; they were subjected to working conditions that were unsafe, tiring, monotonous, and, in industries such as textile mills, very loud and distracting. Injuries were common, and there was no safety net. If the breadwinner was disabled or died, the entire family was immediately impoverished. At the same time, the workplace was becoming much more dangerous due to technological advances, steam- and water-powered equipment, and the newly accelerated pace of work. The end result was carnage. During the Industrial Revolution, the toll of occupational injuries was appalling.

At first, there was little done about unsafe workplaces during the Industrial Revolution because owners subscribed to the assumption that an individual who had an accident was careless. Having an injury was equated with being unsafe—equivalent to behaving badly, in the sense of being irresponsible, like a willful child is a “bad boy” for performing an unsafe act. The worker who took a risk was bad because safety is good and one should know better. Many people still believe

this, and the latent belief underlies the anger reaction sometimes seen in supervisors when one of their workers is injured. On a social level, this punitive concept still persists as the basis for litigation and tort law, which emphasizes single causes attached to individual action and intention. The Anglo-American–Canadian legal system still acts on the general principle that there is a single cause for any accident—a single person to be held responsible to the degree that his or her actions contributed to the injury—and therefore, there should be a single person who pays for the damage. However, this philosophy was not tenable as a social policy for long, because even careful workers were sometimes injured; and it was difficult even then to blame a child worker for his or, in many cases, her own crippling injury.

The “acts of man” idea had created a context in which it was very easy to blame the victim with an unspoken (or spoken) accusation: “it’s your own fault; you are careless or incapable.” This has led to unfortunate self-blame among injured workers and blaming the victim. One injured and presently disabled worker spoke for many when he declared at a conference: “I can’t blame anybody else. It was my own fault that I had this accident. I was careless, I was distracted for 15 seconds and I did something stupid and I know I was dumb.” Such attitudes of self-blaming are ultimately self-defeating. Nobody can work day in and day out in an intense job with stresses all around—distractions, interruptions, and machinery breaking down—and still focus full attention on the task at hand for hours on end. It is beyond human capability. This recognition that there were indeed limits to human capabilities led to the next step in evolution in thinking about safety, which was to replace the emphasis on the individual worker with thinking about systems. But first there was to be one more cycle of blaming the victim.

More benign managers who searched for an explanation were then influenced at the turn of the twentieth century by the new theories of Frederick Winslow Taylor, described in greater detail below. Taylor made a positive contribution to thinking about injury in his mainstream doctrine, which finally turned away from the notion of individual responsibility and the inherent deficiency in the “acts of man” and stimulated thought about social and systems

factors. Some of this new thinking was a reaction against his work. His efforts to achieve mechanical efficiency by workers, with no regard to their own thoughts or emotional engagement with their work, were bitterly resented on grounds that such an approach was dehumanizing and uncaring, and reduced the worker to a trained beast or a mere cog in a machine. Jokes about “efficiency experts” are basically parodies of what Frederick Taylor was doing in the early 1900s. The premise of the joke is always that the efficiency expert is foolishly narrow-minded. The popularity of these jokes in the United States throughout the twentieth century, long after “efficiency experts” disappeared, may be an example of how, through popular culture, workers responded defensively to a perceived threat to the “workingman’s” livelihood and the quality of work life.

Taylor almost single-handedly created the school of thought called “human factors,” later a foundation of ergonomics. He became famous for his time and motion studies, which were applied to assess work efficiency and to search for “the one best way” to do the job. Taylor introduced concepts of efficiency, human capacity, work pacing, and process design that are recognizable today in safety science and in ergonomics. In his mainstream theoretical framework, accidents were errors of perception or reaction time and could be remedied by correct design. Errors were seen as resulting from a mismatch between what the human mechanism was capable of doing and what the physical mechanism that was being operated required in the way of control. By tackling that mismatch, one could reduce the error rate and thereby reduce accidents. This concept of human factors would become enormously productive, once it was channeled productively and after design technology caught up.

During World War I the Industrial Health Fatigue Board in the United Kingdom, influenced by Taylor, adopted the idea that certain workers had an innate a tendency to cause accidents and speculated that 15 percent of workers were involved in 85 percent of accidents. Many managers seized on a corollary (rather than the mainstream message) of this thinking, which posited that while true “accidents” could happen to anyone, there were certain “accident-prone”

individuals who, by virtue of the way their brains were wired or by a quirk of personality, were prone to injury and to repeated accidents. The only effective action that could be taken by management, it was assumed, was to get rid of them. This was the origin of the stereotype of “accident-proneness” that, to some degree, continues today.

The Industrial Fatigue Board also contributed the important and more constructive concept of a utility curve relating productivity and working hours. They determined that the optimal work week for productivity was 50 hours and that work in excess of 60 hours per week led to a sharp drop in productivity and in the quality of the work.

In 1932 the first modern systems theory of safety and injury prevention was formulated by Herbert Heinrich and was quickly embraced by safety officers and safety engineers. This was the “domino theory,” which considered injuries to be the last step in a series of sequential events. The domino theory brought in human elements and recognized that human beings had their limits and needs. The domino theory gave rise to the famous “three Es” formulation, which remains influential today. This formulation recognizes three elements: engineering, education, and enforcement. Engineering was to eliminate the hazard, or guard or provide personal protection against it. Education would be used to provide insight, new behavioral patterns, and training. Enforcement included monitoring, evaluation, and accountability. The domino theory and the three Es put together many elements that had gone before, but together they put engineering in the forward position. Engineering would later assume even more centrality in the ergonomic model.

Interest in the domino theory has been revived by a contemporary safety scientist: Dan Petersen, who incorporated overload, inappropriate responses to the hazard (such as ignoring it or failing to use protection), inappropriate activities, and situational factors into his model of human factors, the “accident/incident” theory. Instruments based on the theory are designed to measure each term in the model, which is then used to explain or predict safety performance and injury risk.

The modern domino theory recognizes several sequential steps, and therefore sources of error, in the evolution of incidents: safety management (associated with training, task design, motivation for safe behavior), safety program management (operation of the safety management system, including information collection and analysis), command or task assignment, system operations (including policies and procedures), operating performance (such as alertness and performance of the operator or design characteristics of the work process), characteristics of the mishap (such as separation of hazards from people), and results (including response, containment, and mitigation).

The domino theory and the three Es had many drawbacks, however. One was that there was no place in this formulation for critical incidents and near misses, which are by far the majority of incidents. The three Es did not guide safety officers to prioritize among hazards, thus leaving them to treat one set of hazards much the same as another. The formulation was very specific to the workplace and did not relate very well to agricultural hazards or to home safety, which limited its applicability in safety science. Finally, it was too dependent on changes in individual behavior. It is much easier to change the specifications on a machine than to change the specifications on a worker. World War II promoted this way of thinking as engineers were faced with the challenge of designing equipment for the military that could be used safely and reliably by new recruits and under extreme conditions.

Since the 1970s, and especially accelerating in the 1980s, safety science and injury prevention have been largely dominated by the approach of epidemiology and public health, which is to look at the problem in a multifactorial way. Safety professionals who look for multiple causes are more likely to see multiple remedies. They may also try to find among several sometimes-unrelated accident trends the single common factor that is amenable to intervention. This approach accepts risks as being inevitable but also manageable, and accommodates the idea of loss, liability, costs, and other related issues that these other systems had separated out and had not considered together.

The multifactorial model arising out of the public-health approach identifies contributions to risk from the task design, the tools used to perform the work, the organizational framework, the local environment, and the characteristics of the worker.

One may conceive of an incident as an unintended release of energy resulting in an injury. William Haddon Jr., a physician and engineer, realized that safety and injury prevention are primarily about controlling the release of kinetic energy. The idea is to channel energy away from the individual and to minimize the amount of energy that can be released in an unintended way. Haddon proposed ten “countermeasures” as general guides to the prevention of injury by controlling the release of energy (see Table 6.3).

The public-health model recognizes that the elements of an incident include the person susceptible to the injury, the hazard that is capable of inducing the injury, and an environment that brings them

**Table 6.3.** Haddon’s Ten Countermeasures

---

1. To prevent the initial marshaling of the form of energy
2. To reduce the amount of energy marshaled
3. To prevent the release of energy
4. To modify the rate of spatial distribution of release of energy from its source
5. To separate in time and space the energy being released from the susceptible structure
6. To separate the energy being released from the susceptible structure by interposition of a material barrier
7. To modify the contact surface, subsurface, or basic structure that can be impacted
8. To strengthen the living or nonliving structure that might be damaged by the energy transfer
9. To move rapidly in detection and evaluation of damage and to counter its continuation and extension
10. To stabilize, promptly treat, and rehabilitate the injured person

both together in the workplace. Bringing together these elements creates a situation that is the context for the incident. A person may decide whether or not to take a risk on the spur of the moment, may not perceive the hazard, may make an appraisal of it that is faulty in some way, may be distracted, or may make a decision in retrospect that was not reasonable. There is a very critical period just before the incident where these things come together into a particular risk-facilitating situation. Most of those will be misses. There may be a near miss—an incident that does not occur—but with a certain probability those situations will come together and the incident event that results in injury occurs. This results in effects that may include injuries, damage to property, financial loss, legal liability, future costs, and overhead costs. Incidents can be prevented, and when they do occur, their effects can be minimized.

The more sophisticated public-health approach is embodied in a formulation called Haddon's Matrix (Table 6.4), which relates the various factors before, during, and after an incident resulting in injury. This matrix provides a more complete framework for identifying the elements that come together in an incident resulting in injury. Haddon's Matrix is an essential tool in modern injury prevention and analysis.

## PRACTICING SAFETY

Ergonomics provides a technology for changing both the hazardous and overly demanding aspects of the environment and the sensory cues and signals that sustain and consolidate behavioral changes.

One can learn a great deal if one can capture information on near misses. This is extremely difficult because near misses are not usually reported. Injuries and serious disabling injuries especially are reported fairly consistently. Minor injuries are not consistently reported, but one may capture most of them. Accidents with only property damage frequently go unreported. Near misses are rarely reported. In fact, usually there are incentives for employees not to mention them at all. There is a great deal we can do with this information. A

risk-management approach to controlling liability and property damage will be likely to affect injury rates also. It has taken industry a long time to recognize this concept in a formal way and to build it into systems of injury control.

Basic principles of injury prevention create a framework for injury or accident prevention (see Table 6.4). One of the most fundamental is that one does not knowingly expose oneself to an unacceptable risk. There are two important considerations in that one statement. The first is embodied in the word “knowingly” and implies an education and training function. If one does not realize the risks one is taking, one may not realize whether or not those risks are unacceptable. The second is the “unacceptable” term, which varies with circumstances. Adults routinely take risks in order to achieve their objectives. Most adults accept a risk of serious injury every time the decision is made to drive an automobile, because the odds of having a serious injury are small compared to the odds of receiving a substantial benefit and because they cannot function without transportation.

Contemporary thinking is therefore moving away from the idea that taking risks is always indefensible. An adult who is out to achieve a purpose is different from a child who takes a risk without knowing the consequences or an adolescent who does not identify with the consequences of an action. In adult life, it is normal to take a risk, and sometimes necessary. Therefore, safety for a child means not taking a risk and following the advice of the parent. Safety for an adult means accepting that a risk can be reasonable in context, but minimizing that risk as best one can.

This concept was explored in a perceptive study by W. A. Harrell (1984) of predictors of farm incidents resulting in injury among members of Unifarm, an agricultural cooperative in Alberta. He found that individual excesses in farm-related injuries related closely to attitudes that farm incidents are inevitable and that farming is a financially risky undertaking. Risk taking was seen to be necessary for profitability, and “cutting corners” was considered necessary for successful farming. Personal characteristics, other than a history of incidents and experience on the job, were much weaker predictors

**Table 6.4.** Haddon's Matrix of Factors Involved in Incidents Resulting in Injury: Examples

Person	Incident	Environment
<i>Pre-incident</i>		
Pre-incident characteristics	Sensorimotor skills	Space, time separation
Sensory acuity	Competing events	Physical factors
Training	Intrinsic hazard	Maintenance
Experience	Judgment	Access control
Stress	Ease of control	Safeguards
Impairment	Kinetic and potential energy	Signs, visual and sensory clues
Distraction	Performance	Legal, regulatory framework
<i>Incident</i>		
Susceptibility to injury	Kinetic energy at contact	Exposure
Positioning	Speed of energy release	Architecture or design
Personal protection	Geometry of contact surfaces	Topography
Reaction time	Containment or dissipation of energy	Barriers and protections
<i>Post-incident</i>		
Immobilization	Protection until first responders arrive	Emergency response
Rehabilitation	Secondary hazards	Medical-care access
Individual treatment	Cleanup	Rehabilitation services available
Psychological adjustment	Presence of an escape or means of access to victim	Retraining services available
Learning ability	Recording devices or forensic evidence of what happened	Workers' compensation and financial support

of risk. These findings seem to be consistent with studies done in industry that suggest that attitude and organizational culture are much stronger predictors of risk than personal characteristics.

Safety is compatible with productivity, however. Managers often think of safety as being something that holds workers back from doing a job—something that gets in the way and that costs money. If one accepts a concept of risk as inevitable in adult daily life, some risks may be required to complete an objective. However, an excessive risk vastly increases the probability of injury, loss, and delay in production, and costs an enormous amount. The culture of the organization and the attitudes of managers may send the message that either risk is tolerated or that needless risk, loss, and liability will not be tolerated.

In recent years, there has been a generational shift in attitudes toward safety among workers. Decades ago, it was common for older workers to be more safety conscious than younger workers, because of their experience and from having witnessed injuries themselves. Today, it is common for younger workers to be more safety conscious because of education and training. Even so, sometimes workers treat serious or unusual hazards in their work with an attitude that seems casual or even foolhardy. There are several likely reasons for this. When one is working with a frightening or unknown hazard, some degree of denial is necessary in order to avoid constant anxiety. One response is to deny the risk and ignore the consequences. This tendency is often reinforced, particularly in groups of young men, by peer pressure to demonstrate “macho” or fearless behavior. Psychologically, this should be seen as a collective form of denial, a maladaptive means of coping with fears by acting them out. Over time, as well, many common hazards become so familiar that they are ignored or considered “acceptable risks,” particularly if there was a tendency to ignore them in the first place. In years past, conspicuous risk-taking behavior was almost a rite of passage into adult working life. On the other hand, older workers sometimes defend their actions with one of two justifications: either they have performed their jobs for so many years without incident and are not about to change, or the damage is already done and they have nothing to lose.

Both (contradictory) justifications are expressions of denial. However, there may be some truth in the idea that a change in routine may result in an event when one has formed habits related to being accustomed to unsafe conditions.

## **WORK ORGANIZATION**

Persuasive studies, such as those conducted by the Finnish Institute of Occupational Health, have identified a number of factors related to work organization that play a major role in predisposing a company toward a higher risk of accidents. There is an inverse correlation between the size of the company and the risk of serious accidents, down to even the smallest enterprises (less than ten workers) in industrial settings. In sectors as different as manufacturing, construction, and transportation, lacerations and other injuries related to moving machinery have been found to be overrepresented among enterprises of all sizes, small and large.

The factors most often associated with serious incidents are related to time pressure, including beating deadlines and saving time on production. Finnish investigators (notably Jorma Saari and Simo Salminen) have suggested that “incautiousness,” by which is meant an unconscious disregard for consequences in an effort to get the work done expeditiously, is a major risk factor. Other factors played a much less important role: wage system (piecework versus hourly wage), “professional pride” (possibly including “showing off” behavior), influence of co-workers, influence of foremen, demands by customers, and misplaced curiosity.

Employees of subcontractors are at special risk. A hugely disproportionate number of incidents related to injury occur among employees of subcontractors in the construction industry: approximately one quarter of injuries in the industry overall, which outweighs experience on the job as a predictor of individual risk. This is probably because subcontracting work exposes the worker to an unfamiliar work site, and because there may be less-satisfactory worker education and enforcement of work practices when a contractor turns the responsibility over to a subcontractor.

One of the eternal questions about work-site safety is: Why is it so difficult to show progress, even though the problem is so apparently simple? The fact is that achieving progress in safety is difficult because changing human behavior is difficult. Change in behavior is not readily accomplished with single-point interventions; the change in behavior must be supported, reinforced, and sustained over time in an environment that does not encourage lapses into familiar old routines. In other words, safety is a profoundly social activity, because changing risk-taking behavior depends on social interactions and social factors such as work organization, not just individual perceptions and motivation. Group behavior cannot be changed effectively by addressing a collection of single persons acting in isolation, and it cannot be influenced by interventions designed to reach them as individuals. Effective safety programs take into account the interaction between people and the social environment, including how work is organized.

Currently, there is a revival of the ideas of Heinrich, applying contemporary ideas of behavioral change and borrowing from management theory on quality improvement. This movement is called “behavior-based safety” and appears promising, but it is very intensive in practice. The ideas could easily be integrated into a broader corporate initiative if, for example, a company were to adopt “Six Sigma” (a management approach for corporate excellence and efficiency) or something similar. Injury prevention is primarily risk management. If one manages injuries, one also potentially manages risk, loss, and costs on an enterprise level.

The relatively new behavioral science emphasis is aligning safety science with the sister professional communities of injury prevention and injury epidemiology, and interaction among these disciplines can only be productive.

## PROGRAMMING SAFETY

Injury prevention can be practiced on several levels. Safety engineering uses a set of tools that follow certain general principles of controlling hazards. Tools available to solve a safety problem include

safety engineering, state-of-the-art ergonomics, behavior engineering, training, and plant design. However, there must also be a framework to use them together effectively. This is where programming is important. An effective safety program applies strategies and tactics that put those tools where they are needed and ensures maximum benefit. This means injury and accident investigation, a monitoring system, risk analysis techniques, worker participation, goal setting, defined objectives in injury reduction, and incentives and recognition for safe workers. Safety approaches and tools should be integrated into a program so that tools, strategies, and tactics are part of a coherent system or program that can be used to ensure ongoing stability and support for injury reduction. These programs should be based on a reasonable injury-causation model and should be appropriate to the population of workers in the workplace. If workers have trouble speaking and reading English, the manager should have a better way to reach them than printed materials in English. Safety measures should be integrated into a comprehensive package so that they are mutually reinforcing; backed by company policy; specific to the hazards in that workplace; clear and explicit; and aimed toward realistic, obtainable objectives that everyone understands. Supervisors should be able to evaluate progress.

An injury-control program has to be reasonable in cost to work effectively. If it is not, management will not be able to sustain it from year to year. Something that is elaborate, expensive, and out of line with the budget—or that the employer can only afford for one year—will result in a flurry of interest that cannot be sustained and ultimately will produce disappointing results.

Anticipating where problems are likely to arise is a fundamental strategy in injury reduction. Indexing systems, an audit procedure to identify the potential for fire and explosion in a workplace, and the Dow index preliminary hazard analysis are some examples of this approach. Hazard and Operability Analysis (universally known as HAZOP) is a popular and effective audit system that can get to be moderately elaborate in seeking early identification of the potential weak spots in a production process, where problems may arise.

Fault-tree analysis is a mathematically sophisticated concept that allows a systematic examination of the process stream and the plant, in order to identify problems before they happen. Unfortunately, these are expensive systems, well out of reach of small enterprises.

The U.S. National Committee for Injury Prevention and Control produced a series of recommendations for policy changes to encourage injury reduction and to motivate employees to develop programs for this purpose. These recommendations are restated in Table 6.5.

**Table 6.5.** General Recommendations of the U.S. National Committee for Injury Prevention and Control, 1989

- 
1. Improved data collection and linkage: fatality investigations, workers' compensation data, inspection records, vital statistics files, healthcare and hospitalization, trauma and other registries
  2. Improved occupational health and safety standards with monitoring of compliance and evaluation of outcome
  3. Criminal prosecution for employees who disregard safety in work practices, and sanctions against those who discriminate against employees who report unsafe working conditions or exercise their right to refuse
  4. Creation of economic incentives with workers' compensation stronger than those that now exist, in order to motivate employees and to shift economic burden onto employees with poorer performance, with evaluation of outcome
  5. Joint labor-management health and safety committees with authority to convert unsafe working conditions
  6. Provision of consultation services to assist employees, funded by workers' compensation assessment surcharges, with evaluation of outcome
  7. Encouragement of shared services and resources for small businesses
  8. Abandonment of incentive awards and contests, because they encourage underreporting and place emphasis on the worker without equally emphasizing the responsibility to provide a safe workplace
-

A critical aspect of programming successful injury prevention, loss control, and risk modification programs is the communication of a consistent, clear, and enforceable message within the organization, stating that safe work practices are the norm and unsafe practices will not be tolerated. Posters have relatively little impact and what impact there is is short-lived. Sustained information campaigns, behavioral (incentive-based) programs, and well-designed training programs each may have a greater or lesser, and often transient, effect on worker behavior, but none of these intervention strategies are of much use in the long term unless there is a feedback process that reinforces the adoption of new values. The most successful approach to injury prevention has been described by Scott Geller in these terms: “safety must *not* be considered a priority in an organization, because priorities can be shifted according to the demands of other priorities. Instead, the ultimate goal of a safety program is to make safety a *value*, associated necessarily with every priority in the organization.”

## RESOURCES

- Friend M. *Fundamentals of Occupational Safety and Health*. 4th ed. Washington, DC: Government Institutes; 2006.
- Wilson L, McCutcheon D. *Industrial Safety and Risk Management*. Edmonton: University of Alberta Press; 2003. [A superb book; it deserves to be a classic. Wilson, the principal author, is an engineer who communicates well with non-engineers.]

## NOTEWORTHY READINGS

- Guidotti TL. Occupational injuries in Alberta: responding to recent trends. *Occup Med*. 1995;45(2):81-88.
- Harrell WA. Factors influencing involvement in farm accidents. *Percept Motor Skills*. 1995;81:592-594.
- Harvey MD. *Theories of Accident Causation*. Edmonton: Alberta Workers' Health, Safety, and Compensation, Government of Alberta; December 1984.
- Kletz TA. *What Went Wrong? Case Histories of Process Plant Disasters*. Houston: Gulf Publishing; 1985. [Classic]
- Rees CD. *Occupational Health and Safety Management: A Practical Approach*. Boca Raton, FL: CRC Press; 2003.

*This page intentionally left blank*

# **7 RISK SCIENCE AND RISK COMMUNICATION SKILLS**

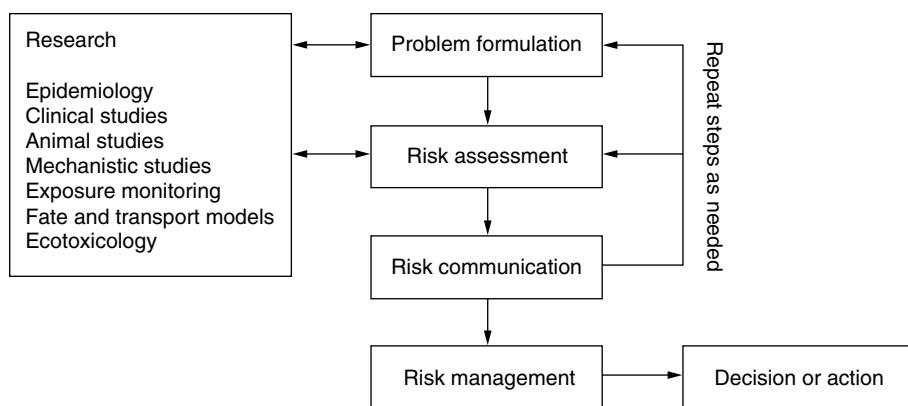
How, exactly, does one figure out how much risk is involved in a particular situation and what to do about it? How do people view the situation in their own lives and as it affects their neighborhood? Once a course of action is decided, how does one communicate the risk message? Collectively, these dimensions of characterizing and managing risk are called “risk science.” “Risk assessment and management,” in the context of hazard identification and control in enterprises and individual workplaces, is discussed in Chapter 14.

Risk science is an emerging discipline that involves the systematic study of risk and uncertainty, treating it as a formal object of study. In this context risk is the probability of an adverse event or consequence, such as an injury in the workplace or developing cancer after exposure to a carcinogenic chemical. For occupational and environmental medicine, it represents a structured approach to the evaluation and management of risks, primarily as they arise from environmental or occupational hazards, consumer products, and lifestyle. In public health generally, risk science is rapidly emerging as a “basic science” that, together with epidemiology, provides a foundation for analysis, regulation, and planning. It can also be seen as an intermediate discipline that brings the knowledge of toxicology, epidemiology, and other relevant

fields to bear on the evaluation of health-related problems and the steps required for their solution. Although risk science is most often used for consequences involving human health, it is also used for ecological impact, psychological impact, and anticipating costs.

The basic tools of risk science have broader application and can be used wherever there is an element of unknown or uncertain risk. Applied to engineering, risk science is primarily concerned with the possible reasons for failure in structures or processes. Applied to financial markets, it is primarily concerned with predicting the likelihood and degree of volatility and managing investments to reduce loss and maximize gain. Applied to insurance, it is primarily concerned with reducing exposure to loss and identifying interventions that can save property or lives and so has a close relationship with the professional field of loss management.

The basic steps in evaluating the risk of an occupational, environmental, consumer, or lifestyle hazard are obvious intuitively. They are given formal names, such as those designated by the U.S. Environmental Protection Agency (EPA) (Figure 7.1). First, the hazard must be identified (“hazard identification”). Then the degree of risk arising from the hazard must be characterized (“risk characterization,” sometimes called “risk analysis” or “consequence assessment”); logically this would be done by determining how many people are



**Figure 7.1.** Risk assessment and risk management paradigm developed by Health Canada.

exposed and who they are (“exposure assessment”), and then determining how the hazard is most likely to affect them (“dose- or exposure-response”). Then there must be a phase of determining what, if anything, can be done about the problem (“options development”); a decision would be made on the best course of action (at this point not a scientific decision but one taking into account values, costs, politics, and social acceptability), and this would lead to concrete implementation of steps to deal with the issue (“risk management”), which inevitably include measures of communication and education (“risk communication”). Different countries, agencies, and organizations have called these steps by different names and organized them differently, but the basic logic is usually followed in this sequence.

The constituent subdisciplines of risk science will be discussed in turn, followed by a discussion of standards and guidelines:

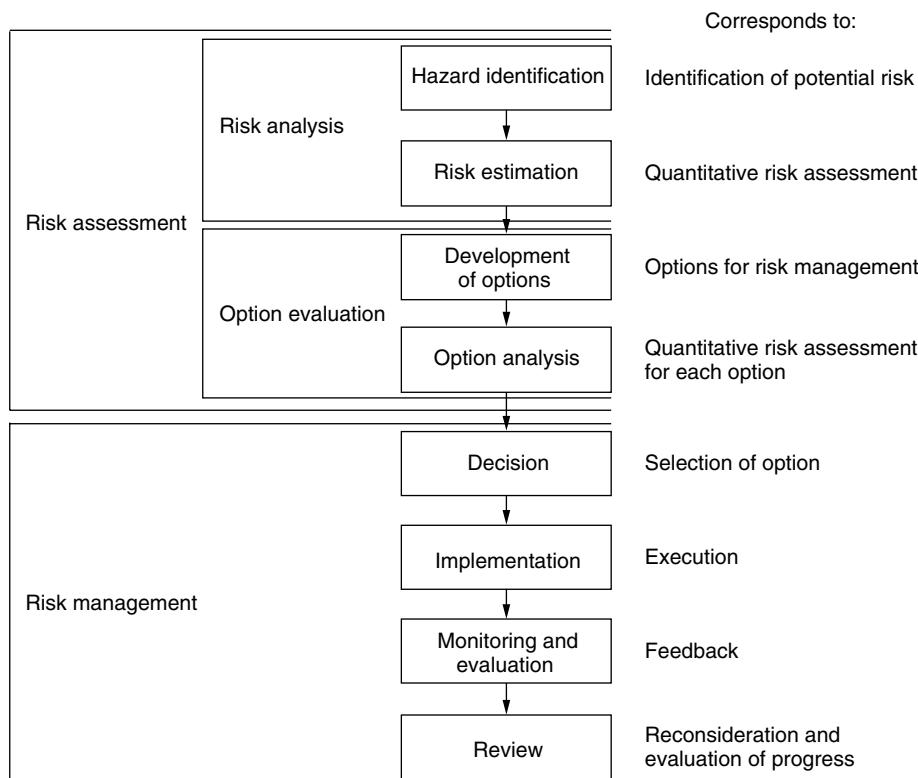
- Risk analysis, the evaluation of the risk presented by a particular hazard
- Risk assessment, the evaluation of a risk problem, taking into account the context of the hazard and the population exposed, which may be qualitative or quantitative
- Risk perception, the way communities and individuals perceive risk-related problems and the determinants of their concepts, attitudes, and beliefs toward risk in general and toward specific risk problems
- Risk communication, a systematic approach to conveying information on risk-related problems and providing the knowledge to people that allows them to make decisions about specific problems
- Risk management, the systematic approach to interventions and options for managing, ameliorating, or reducing risk
- Risk comparison, a structured way of comparing degrees or classes of risk so that priorities can be set

The terms used in this chapter are in common use in environmental sciences and in regulatory affairs. Terminology tends to vary

in other fields. For example, risk management means regulatory options in government, liability management in law, and reduction of financial and liability exposure in economics and business.

## RISK ANALYSIS

Risk analysis is an overarching term for the quantification and systematic evaluation of risk. It is sometimes used as a synonym for risk science, and it is also used to refer to an early stage in the process of risk assessment, as by Health Canada (Figure 7.2).



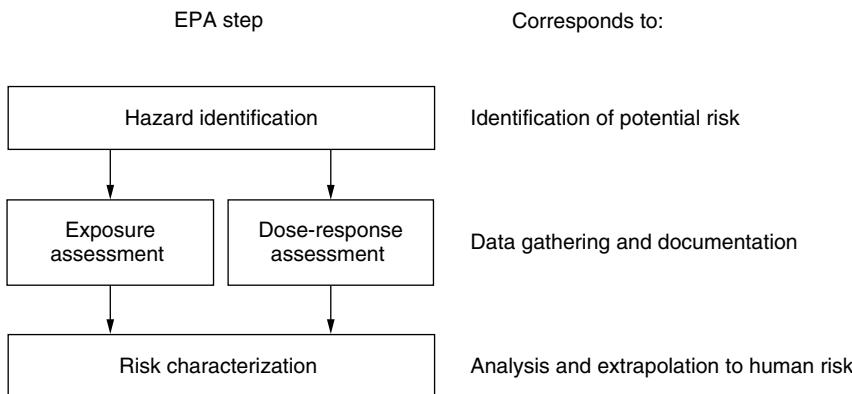
**Figure 7.2.** Risk assessment and risk management paradigm used by Health Canada.

Different agencies, countries, and organizations apply different names and put the pieces together a little differently. In Canada (see Figure 7.2), the development and evaluation of risk management options are also considered parts of risk assessment, as steps coming before the decision process. On the other hand, the EPA does not separate out options analysis from the decision process, which falls under risk management (see Figure 7.1 and Figure 7.3). Risk analysis, as a general term, implies the overall process rather than any one part of it.

Common to all notions of risk analysis is the idea that uncertainty can be anticipated and quantified and that the laws of probability and reasonable estimation can, within bounds, be used to create rational models for making decisions. As such, risk assessment, in particular, has become the principal tool of government regulatory agencies in environmental and occupational health. The remainder of this chapter will discuss risk science from the general point of view of regulation and regulatory agencies.

Risk itself is a conceptual abstraction. A risk is a probability that something, usually bad in this context, will happen. It is not an intrinsic property of anything or a tangible object. Rather, it is a description of the behavior of a usually complex system (a hazard, the opportunity to encounter it, the level of exposure that is likely, and the population that encounters it, including susceptible groups) that is generally based on what occurred in the past. Risk therefore represents the characteristics of a situation at a given time and if the underlying factors change, then the risk changes.

All society is, in a sense, about mitigating risk. People probably started to live in communities in the first place for protection, in order to help one another (if only to bring down the odd mammoth), to reduce the risk of not finding a mate, to trade and therefore reduce the risk of shortages, and to communicate information that helps in deciding how best to manage risk. As society becomes more affluent and major risks of survival are controlled, other risks become more prominent and disturbing and so become social issues.



**Figure 7.3.** Risk assessment paradigm used by the U.S. Environmental Protection Agency, in detail.

## RISK ASSESSMENT

Risk assessment is the evaluation of the likelihood of an adverse effect occurring as a consequence of some exposure in a population. A prominent risk expert, Vince Covello, has written that risk assessment “is a systematic process for describing and quantifying the risks associated with hazardous substances, action or events.”

“Qualitative risk assessment” is a matter of informal assessment after a review of the evidence; it is the sort of informed judgment that occupational and environmental physicians make every day. Qualitative risk assessment is the basic approach of public health practice, where judgment and experience are used to identify and prioritize environmental and occupational hazards. “Quantitative risk assessment” is the application of a formal model that draws on principles and methods of toxicology and epidemiology. It is a means of documenting assumptions, enforcing transparency, making steps in the evaluation of risk highly explicit and consistent from one regulatory problem to another, and identifying gaps in knowledge that need to be addressed in order for a reasonable estimation of risk to be made.

Quantitative risk assessment (hereafter just called risk assessment) is particularly valuable in allowing comparison of various risks. Although estimates of risk for any one exposure often span an order

of magnitude because of cumulative uncertainties in several terms, the risks for various exposures may be ranked by magnitude for comparison. This is exceedingly important in proposing new occupational or environmental standards, in selecting which exposures are significant from the public health point of view and which are less important for immediate attention, and in explaining risks to the public.

Risk assessment is a conceptually difficult process that draws on science but is not science in the formal sense; it does not generate falsifiable theories and hypotheses and evaluate them against empirical evidence. Rather, it builds a model based on what is known to estimate the level or range of risk likely to arise from a set of circumstances such as exposure to a chemical hazard. Risk assessment is a means of introducing rigor into the evaluation of a possible threat to the public health or interest.

The modern age of risk assessment began in 1983 with the release of the so-called Red Book by the National Research Council. This blueprint for conducting quantitative risk analysis was prepared for the guidance of the Environmental Protection Agency (EPA) and soon became the template for other federal agencies engaged in regulating consumer and environmental risks, including and especially the U.S. Food and Drug Administration. The Red Book is currently undergoing revision. Figure 7.3 outlines the current EPA paradigm for risk assessment.

The most critical early exercise in risk assessment was the saccharin study conducted by the National Academy of Sciences in 1978. This study of the potential for induction of bladder cancers following intake of saccharin in foods was a paradigm for the difficult problems that were to come. In order to perform the study, a number of techniques had to be pioneered by the Academy and a great deal was learned from the experience. As a result of animal toxicity studies performed in Canada in 1977, it was found that rats fed high doses of the noncaloric sweetener saccharin developed bladder cancer at a frequency much greater than expected. This led to a ban on saccharin as an artificial sweetener by the U.S. Food and Drug Administration acting under the Delaney clause (a 1958 amendment to the

Food, Drug, and Cosmetic Act of 1938), which required it to withdraw from human consumption any artificial food additive demonstrated to be carcinogenic in animals. The U.S. Congress overruled the ban and handed the problem to the National Academy of Sciences to be studied. The assessment largely hinged on the potential risk to individuals consuming large quantities of saccharin in soft drinks, a particular concern for diabetics. At the time, aspartame was not available and the cyclamates had been withdrawn earlier because of similar concerns regarding bladder cancer. (Cyclamates were later reevaluated, vindicated, and returned to the marketplace.) In the end, when the report was released in 1979, the Academy came to the conclusion that between 15 and 15,000 additional cases of bladder cancer a year might be caused by the intake of saccharin, which they concluded to be a definite promoter (or co-carcinogen) but a questionable or very weak initiator (primary carcinogen) of cancer. That this set of estimates spanned three orders of magnitude reflected what was obviously a difficult assessment problem. However, the saccharin study defined the issue well and provided a framework for resolving the issue in Congress. Saccharin was allowed to remain in the market after it had been convincingly demonstrated to be such a weak potential carcinogen that its continued use was considered to be acceptable.

In 1980 a decision of the U.S. Supreme Court imposed a requirement on the Occupational Safety and Health Administration to apply risk assessment in the formulation of standards, in a case involving regulation of exposure to benzene. Within the U.S. federal government, oversight of the regulatory process has been progressively concentrated in the Office of Management and Budget, which, in 2007 and 2008, placed progressively more stringent requirements for documenting the risk assessment process on agencies. This was interpreted by many as a ploy to slow down or prevent further regulatory actions and by others as a means to ensure that risk assessment remained grounded on sound scientific principles.

The intellectual history of risk assessment has been driven by regulation and characterized by the search for a framework that can

accommodate all forms of hazard in a consistent manner. In the field of occupational and environmental health, risk assessment has been closely linked to the evaluation of cancer risk because the most common application of the approach at first was to the numerous issues in which carcinogenicity was the major concern. The principles of cancer epidemiology lend themselves particularly well to risk assessment, being a much-studied, stochastic process (see Chapter 1). The EPA soon developed a parallel approach to the greater number of hazards that follow an exposure-response relationship for magnitude of effect, later reconciling the two into a common framework. Over the last decade, the EPA has also developed a template or paradigm for microbial risk assessment, treating the probability of infection (another stochastic process) and complications within a framework for the estimation of risk for biological hazards.

Although the EPA took the lead in the United States, the development of risk assessment has been a global, cooperative venture. The U.S. Food and Drug Administration, World Health Organization, Health Canada, regulatory agencies in other countries, and private organizations (notably the International Life Sciences Institute) have all contributed to the intellectual development of the field as well as its application to various problems. Perhaps the most active venue for risk assessment today is the European Union, which in the view of many has become the chief regulatory driver for international trade and product safety and the world leader in occupational health and safety regulation.

The assessment of risk is the first step in the management of risk. The two functions are always kept conceptually separate, however, because of the danger of influencing the assessment of risk to reach a preconceived outcome. When risk management and risk assessment are combined, it is too easy to manipulate the risk assessment to support a pre-established position or conclusion. The assessment of risk for purposes of government regulation figures in the agendas of many stakeholders, who are concerned about such matters as cost, administrative efficiency, legal liability, and the political implications

of various control strategies that may be imposed on a product or industry. Charges of conflict of interest are common. There is therefore a real danger when the same people who are invested with the responsibility for regulation perform the risk assessment. There is a long-standing mordant joke in the field in which a decision-maker asks the risk assessor what the risk is for a particular drug or pollutant, and the risk assessor asks, “What level of risk do you want?” It is easy to manipulate the outcome by making different assumptions, tweaking the model, and being selective about which data are used. It is preferable to have an independent risk assessment group that is firmly grounded in epidemiological and toxicological technique do the work.

This potential for conflict plagued the U.S. Atomic Energy Commission before its regulatory arm became the Nuclear Regulatory Commission and its research functions were reassigned to the Department of Energy. The issue also has surfaced from time to time at the EPA. Avoiding conflict was a major reason for separating the National Institute for Occupational Safety and Health from the Occupational Safety and Health Administration when the two were created by the Occupational Health and Safety Act in 1970, although some might say that the division of functions resulted in two agencies each too weak for their purpose.

### **Hazard Identification and Characterization**

In risk assessment, universally, the hazard is first identified and then characterized with respect to how it affects people and how many and what kind of people are affected. These steps are called, by the EPA, “hazard characterization,” “dose-response assessment,” and “exposure assessment,” respectively. These early steps produce the critical information needed to take the next critical step—“risk characterization,” in which the level of risk to the community or to a single person is calculated.

Risk assessment has its own terminology, which is not always consistent in usage among those within or with those outside the field

(see Table 7.1). For example, the term “risk assessment” is used in regulations of the Department of Housing and Urban Development to refer to the evaluation of homes for the presence of a lead hazard and is sometimes used for identifying health risks for individual patients. Risk assessment in financial circles means assessing the likelihood of losing money and hinges on the volatility of stocks. In insurance, it means estimating the risk of loss due to disaster or excessive claims. The terminology of risk assessment in regulatory policy and public health, which came late to this discipline’s vocabulary, should be understood as comprising terms of art, with their own meaning within this context.

The most generally accepted standardized vocabulary comes from the National Research Council. The general approach to human health risk assessment at the EPA is described in Figure 7.1. The first step in risk assessment is hazard identification, when the intrinsic hazard (such as toxicity, safety, failure rate, risk of loss, or other measure) is first identified. This is usually the result of scientific research, media attention, or a dramatic event. The next two steps occur in tandem: dose-response assessment, discussed below and in Chapter 2, and exposure assessment, described in Chapter 4 and briefly below.

Dose- or (conceptually preferable) exposure-response assessment defines the level of hazard associated with a particular exposure. Frequently, exposure-response relationships must be extrapolated to humans from animal studies because epidemiologic studies may be lacking or provide insufficient information in the exposure range of concern. The quantification of hazard in this way defines an intrinsic property of the exposure for a given end point. For example, exposure to lead may be viewed in terms of potential cancer risk, fertility effects, central nervous system effects, hematologic effects, and systemic effects. As lead is increasingly well controlled in the environment, exposure decreases and body burden drops to levels at which the more gross effects become rare. The risk of more subtle outcomes, such as neurobehavioral effects and cancer, emerges as an increasingly important driver of public policy, particularly for susceptible groups such as young children.

**Table 7.1.** Definitions of Common Terms of Art in Risk Science

---

**Exposure.** Usually, shorthand for the magnitude of contact between a person and a hazard that results in an influence on that person, such as absorption of a xenobiotic, release of energy in the person's immediate proximity, perception of a stressful work environment, or direct contact with a contagious microbial hazard. An exposure occurs only if there is a pathway to and into (or onto) the body. If the hazard does not reach the person, there is no exposure. Also often used as a synonym for "hazard."

**Hazard.** The proximate cause of an adverse event, such as a toxic chemical, an unsafe work environment, a psychologically stressful environment, or a biological threat. Also, the intrinsic property of such a cause that is capable of producing injury after exposure, such as toxicity, energy released, stress, infectivity, allergenicity, etc. For a chemical, hazard is described by the exposure-response relationship.

**Opportunity for exposure.** The probability of exposure to the hazard occurring under all or certain circumstances. Various groups and subgroups in the population, and jobs in a workplace, may (and usually do) have very different opportunities for exposure.

**Population.** In risk science, a shorthand term for the people who are at risk and the cluster of characteristics—demographic, social, and biological—that determine their risk. When used alone, the term usually refers to the population of individuals actually exposed to a hazard. There may be great differences in opportunity for exposure and magnitude of exposure. Characterizing the population at risk therefore requires taking into account the distribution of characteristics (such as employment, consumption of a product, or residence in a certain area) that determine exposure, those factors that determine vulnerability (such as environmental justice issues or lack of means to avoid exposure), and those factors conferring susceptibility (infancy, advanced age, pregnancy, medical conditions, etc.).

**Risk.** Applied to environmental and occupational risk science, the probability of an adverse event such as a disease, injury, disability, or death. Risk is a function of the intrinsic properties or influence of the hazard on people (toxicity, infectivity, allergenicity, energy released, stress created, and so forth), the opportunity for exposure, the magnitude of exposure, and the different populations, which can be characterized demographically and by vulnerability. Risk in other contexts may relate to animal health, ecosystem effects, liability, financial loss, and other consequences.

(Continued)

**Table 7.1.** (Continued)

---

*Risk analysis.* The systematic study of risk, usually used as a broad term for the elements of risk assessment and risk management.

*Risk anticipation.* An emerging field within risk analysis that seeks to predict which emerging hazards or situations will become regulatory problems or public issues and therefore require early attention and higher priority.

*Risk assessment.* A systematic approach to the evaluation of risk. It may be *qualitative*—based on judgment, knowledge, and experience, which is the norm in public health practice—or *quantitative*, based on a rigorous analysis, which is the norm in regulatory affairs. Quantitative risk assessment is built on principles of epidemiology and toxicology and uses data from these fields to provide information in a framework that is useful for making decisions.

*Risk communication.* A systematic approach to communicating messages about risk and uncertainty. As a field, risk communication draws heavily on cognitive psychology, communications theory, and linguistics. It is also a practical skill required of anyone with a professional interest or duty in talking to people about risk, whether the general public or decision-makers.

*Risk comparison.* A systematic approach to setting priorities and studying various risks to assess their relative importance to the community. The best-known approach to risk comparison was a project undertaken by the EPA in the 1980s that featured repeated rounds of extensive community consultation and interaction with experts. It is widely thought to be unwieldy and impractical for routine use.

*Risk determination.* (Canada) The assessment of the magnitude of risk and options for dealing with it.

*Risk perception.* The description or systematic study of the way individuals or groups of people view threats, risks, and uncertainty. Dimensions include how acceptable people may find a given risk, how threatening it seems to them, the implications they perceive for themselves or loved ones, and the implications they perceive for their community. As a field of study, risk perception draws heavily on cognitive psychology.

*Risk science.* A scientific approach to all aspects of managing risk and uncertainty, geared toward deriving general principles and consistent approaches that can be applied in many situations.

---

Exposure assessment, in this context, seeks to describe how often people are exposed, who is exposed (with a special emphasis on susceptible subpopulations), and how intensively they are exposed. Whatever the degree of hazard, characterizing the opportunity for exposure within a population is just as important as characterizing the level of exposure when it occurs. If an exposure does not occur, it makes no difference how intense it would have been. On the other hand, when a relatively low-level hazard becomes extremely widespread in the environment and comes to expose a large number of people, the resulting public health hazard may be significant simply because the small effect is multiplied by the huge number of people exposed.

Another factor subsumed (and sometimes lost) in the exposure assessment step is the population at risk. It is unlikely that all of the population will be exposed to a particular hazard at a uniform concentration. The population at risk is the total number of individuals who have the opportunity to be exposed. Within the exposed population there are groups more likely to be affected than others, typically the very young, the very old, women in pregnancy, and those with a specific susceptibility state, such as asthma or inborn errors of metabolism. A complete exposure assessment requires compilation of the most accurate information on opportunity for exposure, level of exposure, and population at risk. Where information is not available, it is necessary to make supportable, evidence-based, transparent, and reasonable assumptions. This may involve biologically appropriate models, especially to extrapolate risk estimates to lower exposure levels. All this is normally done within conventional bounds of statistical uncertainty to yield upper and lower confidence intervals.

## Risk Characterization

In risk characterization, or “consequence assessment,” one puts all of the information together and derives an estimate of risk, which is calculated with an accompanying band of uncertainty, usually a confidence interval or a range. Among the estimates that may be provided are the most likely outcome (corresponding to a point estimate in

epidemiology), the worst-case scenario, and the 95th percentile of the confidence interval (upper confidence limit or upper bound, corresponding to a condition less than the worst case but still extreme). For regulatory purposes, the worst-case or the upper bound is the estimate typically used. Risk management based on estimates that are assumed to overstate the risk is more protective.

Epidemiological data are preferred for risk characterization because they reflect human experience. When they are not available, animal data must be used to estimate human risk, with suitable adjustments (for example, for body surface area) made based on knowledge of the limitations and differences in biology, if known. It is customary to consider human beings to be as susceptible to a toxic agent as the most susceptible known animal species unless there is a known difference in the way that animals respond compared to human beings. For example, rodent species behave quite differently after exposure to any of a number of hydrocarbon agents, such as gasoline, than human beings. They exhibit an effect on the liver (peroxysome proliferation) that does not occur in humans. For this reason, extrapolations from animals are not used for cancer risk assessment when they involve this mechanism.

It is now clear that, in general, human beings are not as sensitive to many and perhaps most chemicals (such as dioxins) as are other species. Therefore, considering humans to be as sensitive as the most sensitive animal species implicitly introduces an additional degree of safety in the later calculations.

Quantitative risk assessment is predicated on a number of assumptions, which are summarized in Table 7.2. One of the important assumptions underlying quantitative risk assessment is that animal data can be used, selectively, to anticipate possible risks to humans. In order to validate this grand assumption, a great many specific assumptions are required regarding extrapolation of animal studies to humans. In general, animal toxicity studies are performed at high exposure levels on the most sensitive species available. As a practical matter, dosage (crudely reflecting integrated exposure levels) is adjusted by body surface area to obtain at least rough comparability between animal

**Table 7.2.** General Assumptions in Risk Assessment

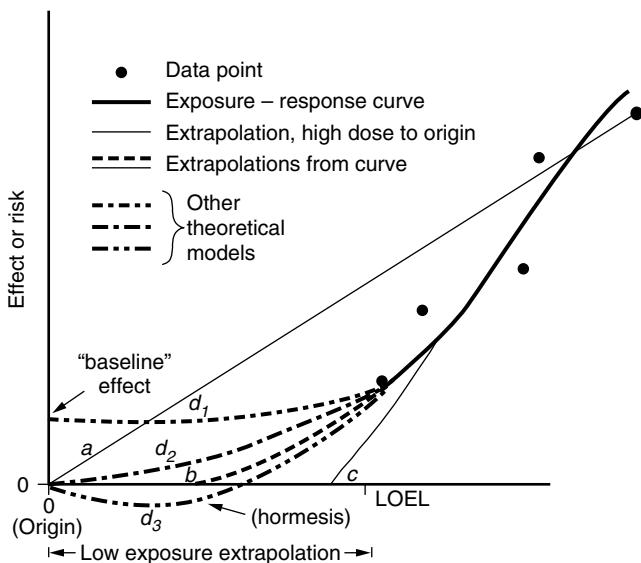
---

1. Human data obtained through epidemiological studies are preferable to animal data obtained through experimental studies as the basis for exposure-response relationships used to calculate risk estimates.
  2. The use of human data must take into consideration the quality of the data that are available and intrinsic uncertainties in the data.
  3. Taking the upper 95 percent confidence limit in exposure-response estimates and applying uncertainty factors corrects for unknown (nonstatistical) uncertainty and increases the margin of safety. For that reason, these procedures are standard for risk assessment as well as risk management options.
  4. Human beings are to be considered as sensitive as the most sensitive known animal species.
  5. Corrected for body surface area, dosages can be compared among species and extrapolated from susceptible animals to humans as a first approximation.
  6. Animal species are generally satisfactory surrogates for human beings in assessing toxicity unless there is a known difference in toxicokinetics, biotransformation, or mechanisms of toxicity. When there is evidence to suggest otherwise, intensive study of interspecies differences to explain the dissimilarity is required.
  7. Lifetime, chronic, or subchronic exposure studies in animals are considered to roughly correspond biologically to lifetime exposure in human beings for purposes of identifying chronic effects such as cancer. It is acknowledged that experimental animal species (usually rodents) have much shorter lives and require higher levels of dosing to induce cancer within this limited time period, during which their bodies generally undergo pathophysiologic changes such as carcinogenesis more rapidly than would be expected in human beings.
  8. A linear, non-threshold model extrapolating from high exposure levels tends to be the best upper-limit estimate for cancer risk when extrapolated at low doses and therefore provides an intrinsically conservative risk estimate.
  9. Hormesis (the theory that exposure-response curves may pass through a phase in which low-level exposures may be salutary rather than harmful, because repair mechanisms are stimulated) is too uncertain to be used as a basis for risk assessment. If it occurs, such an effect is probably too variable for different outcomes to be factored into a workable model.
-

species and humans. This is often an uncertain exercise, of questionable relevance to exposure in real life, since levels of a carcinogen have to be high in animal studies to result in cancer within the short lifetime of the animal. An entire U.S. federal research establishment, the National Center for Toxicological Research, and the National Toxicology Program exist largely to test assumptions regarding the extrapolation of animal studies to lower doses and comparative toxicology and to apply them to specific critical chemical hazards, respectively.

To construct a theoretical exposure-response relationship for extrapolation to humans, some models must be assumed for low levels of exposure when the exact shape of the exposure-response curve is uncertain, as it almost always is. Figure 7.4 shows an exposure-response relationship based on limited experimental data and the various options for modeling extrapolation to low exposure levels (linear extrapolation, extrapolation from the slope, quadratic, linear-quadratic, disappearance into a “baseline” effect, and “hormesis,” the doctrine that very low levels of exposure induce adaptive or reparative effects that may be beneficial). These models were derived from experience with radiation and cancer risk, which is the best body of empirical evidence for low-level exposure and cancer, although still incomplete. The most appropriate model is almost never known and so is decided based on arguments from principle and by analogy. Extrapolations are therefore very controversial and depend on a number of assumptions that remain unproven.

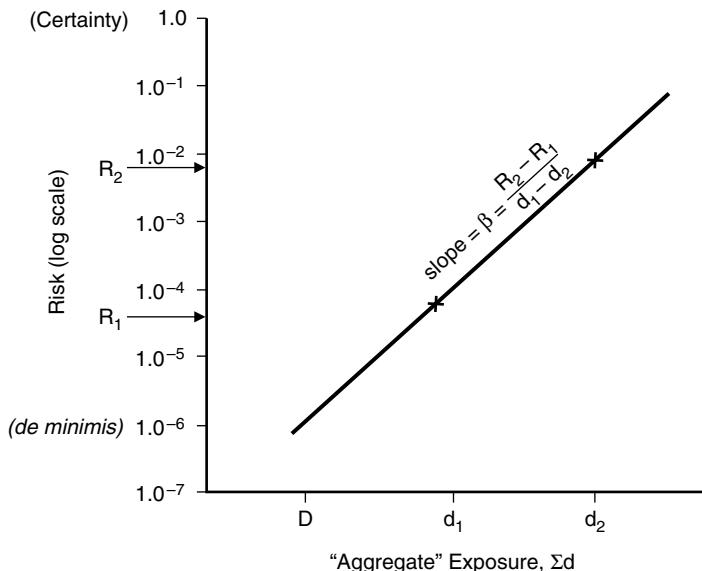
In general, for regulatory purposes involving cancer and other very serious health outcomes, a linear extrapolation (on a regular plot, not a log plot) of the exposure-response relationship from the highest dose to the origin (no exposure, no effect) is used, rather than a guess regarding the shape of the curve or an extrapolation from the slope of the curve or an assumption about where a threshold might lie. Linear extrapolation to the origin produces the most conservative assumption based on the evidence and provides the greatest safety margin. (See Figure 7.4.) The actual shape of the curve at low exposure levels is rarely known with accuracy and is often hotly debated. For example, some toxicologists believe that at very low levels of exposure, DNA repair mechanisms are stimulated out of proportion to new genotoxic events, so that



**Figure 7.4.** Extrapolations for exposure-response relationships at low doses. (Key:  $a$  = linear extrapolation to origin, usually done for cancer risk;  $b$  = threshold of response, usually not known;  $c$  = linear extrapolation from exposure-response curve, which carries a risk of underestimating toxicity at low exposure levels;  $d_1$  = alternative complex model that assumes a baseline effect or risk;  $d_2$  = alternative complex model that assumes a curvilinear exposure-response relationship intersecting the origin;  $d_3$  = alternative complex model that assumes hormesis, a controversial theory proposing that at very low levels, exposure to carcinogens or radiation may induce repair mechanisms that actually lower the baseline response.)

very low exposure acts paradoxically to reduce the cancer risk below baseline. This idea, called “hormesis,” is sometimes generalized to other exposures and effects. Many scientists do not agree and believe that the more likely exposure-response relationship at low levels is a curvilinear relationship that either ends at origin or at some irreducible baseline risk. Almost every risk assessment methodologist would agree, however, that extrapolating from the linear (middle) segment of a sigmoid curve back to the  $x$ -axis yields a figure that is too high.

Once linearity is adopted as a working model, risk extrapolation in principle follows a general, rather simple mathematical idea, illustrated in Figure 7.5. One assumes that a more or less linear relation-



**Figure 7.5.** Paradigm for social dimensions of risk management.

ship can be obtained at low levels of exposure (usually on a log scale, since exposure-response relationships are curved and best represented on a log plot at their lower range). The line that represents the exposure-response relationship has slope  $B$ . A known or extrapolated effect or risk,  $R$ , is associated with a level of dose or exposure,  $d$ . The slope of the line represents the change in the magnitude of the effect (for a gradient or proportional effect) or the probability of an outcome (for a stochastic effect), for a given difference in dose. The slope  $B$  is therefore equal to the change in  $R$  associated with the change in  $d$  over the range of the curve that matters. In practice, the upper limit of the 95 percent confidence interval is usually used for  $R$ . Based on the exposure assessment, an estimate of cumulative or "aggregate" dose,  $\Sigma d$ , is made for a lifetime of exposure. (An assumption of 70 years is typical for the EPA. The EPA prefers the term "aggregate" to avoid confusion with cumulative body burden. The xenobiotic does not have to accumulate in the body, although the example is

simpler if it does. The calculation is based on repeated exposures and the cumulative effect of the exposure events, not retention of the xenobiotic in the body.) The risk,  $R$ , is therefore equal to  $B \Sigma d$ . For carcinogens, the EPA has adopted the regulatory philosophy that a risk of one in a million (either a risk of one cancer in the lifetimes of one million people, or a one-in-a-million chance of developing cancer for an individual) constitutes a *de minimis* risk, low enough to be acceptable to society. If one wants to determine the level of exposure that would be considered “safe” under those circumstances, say  $D$ , one would calculate  $R = 10^{-6} = B \Sigma d$ , so that  $D = 10^{-6}/B$ . A level of exposure ten times greater, associated with a risk of  $10^{-5}$ , would be unacceptable and grounds for regulatory action.

Obviously this is a simplistic example. There are many refinements and elaborations on this basic idea. Different agencies use risk assessment methodologies that fit their particular mandate and regulatory framework. There are, for example, many ways of expressing the slope. For cancer-driven risk assessments, the EPA expresses the slope in terms of the lifetime risk per unit dose in mg/kg aggregate intake, a value called  $q_i^*$  (pronounced “Q star”). Calculated  $q_i^*$  values for known and suspected carcinogens can be found in the EPA’s IRIS database. The known (highest or most certain) dose that corresponds to *de minimis* or absent risk is called the “reference dose” (RfD) by the EPA. The RfD differs from the NOAEL statistic (see Chapter 2) because it is calculated, not observed, and uncertainties are factored in.

The EPA most often applies risk assessment to determine an exposure standard. The Food and Drug Administration uses risk assessment to derive tolerable Total Daily Intake levels for food residues. Health Canada uses Acceptable Daily Intake levels for contaminants and pollutants alike. The EPA and the Agency for Toxic Substances and Disease Registry also use risk assessment to determine the level of risk associated with contaminated sites.

In practice, there may be multiple outcomes (although cancer risk almost always forces the estimate for regulatory purposes when a carcinogen is involved) and there may be reasons to assume a curvilinear exposure-response relationship that is more complicated. There

may be different risk estimates for different groups (such as children), there may be allowances in the assumptions regarding cumulative effect based on the toxicokinetic behavior of the xenobiotic, and one or more uncertainty factors may be applied. The models that are used to characterize risk are subject to validation and intensive study, which may involve sophisticated statistical methods such as Monte Carlo analysis, or sensitivity analysis to determine the robustness of the estimates with changing assumptions.

Risk assessment practitioners have developed the set of reasonably consistent assumptions to ensure consistency and conservatism in the extrapolation of risk that are summarized in Table 7.2. It must be recognized that these assumptions are not scientific in the sense that they reflect actual conditions; rather, they are explicit rules for calculating risk to ensure an unknown margin of safety that is wide as possible and to keep the probability of an underestimate of risk to a minimum. Such assumptions are said to be “risk averse” because their net effect is to minimize risk when translated into options for risk management.

### ***Limitations of Quantitative Risk Assessment***

Quantitative risk assessment, as valuable as it is for clarifying public policy issues, presents a number of problems. Some of these are outlined in Table 7.3.

Critics of risk assessment often perceive it as stretching science to provide a number for purposes of political convenience. It is more accurate to say that risk assessment picks up where science leaves off and packages the information in a form that can be used to make decisions. Risk assessment expresses its findings in terms that are more immediately applicable to public policy decisions than using the traditional terms of epidemiology or toxicology. For example, risk estimates in the traditional form of relative risk or attributable risk in epidemiology (see Chapter 3) provide a useful scientific measure in epidemiologic terms but are of very little help in setting priorities or providing guidance for regulation. The same information is required but in a different expression.

**Table 7.3.** Some Problems of Risk Assessment

- 
- Quantification (conversion into deaths, cases, lives, costs, or probabilities)
  - Comparability (value of human life, trade-offs among outcomes)
  - Exposure-response relationships and low-exposure extrapolation
  - Quantifying the risk of rare catastrophic events
  - Very wide uncertainty bands
  - Taken out of context, single number generated by the risk assessment is misleading
- 

Risk assessment certainly has inherent limitations, but its essence is handling uncertainty and making comparisons of risk, rough as they may be, so that reasonable decisions can be made. The underlying logic of risk assessment rests on the same appreciation of uncertainty that underlies biostatistics but cannot be as accurate in its estimation. Risk assessment is neither accurate nor precise but it provides a working estimate that can be used to support decisions. Rather than telling decision-makers the score, risk assessment tells the decision-makers in what ballpark they are playing.

Generally, the most convenient measure on which to scale health effects is in terms of human lives. It is very difficult to quantify issues of cost, benefit, or risk. There have been efforts to quantify the economic impact associated with health risk in terms of the monetary value of human life. Agencies that have tried to valuate (the term of art in economics) human life have encountered a hostile public reaction. This was the experience of the EPA in 2008 (and of the Consumer Product Safety Commission before it) when it proposed the revaluation of a generic human life for purposes of assessing the costs and benefits of regulation. The EPA proposed reducing the value from about \$8 million to about \$7 million but was criticized on the ground that this move changed the economics of air pollution regulation to make the case for control less favorable. In fact, other economic analyses commonly place a lower value on the average life—as low as \$129,000 for a life of “quality”—as calculated by

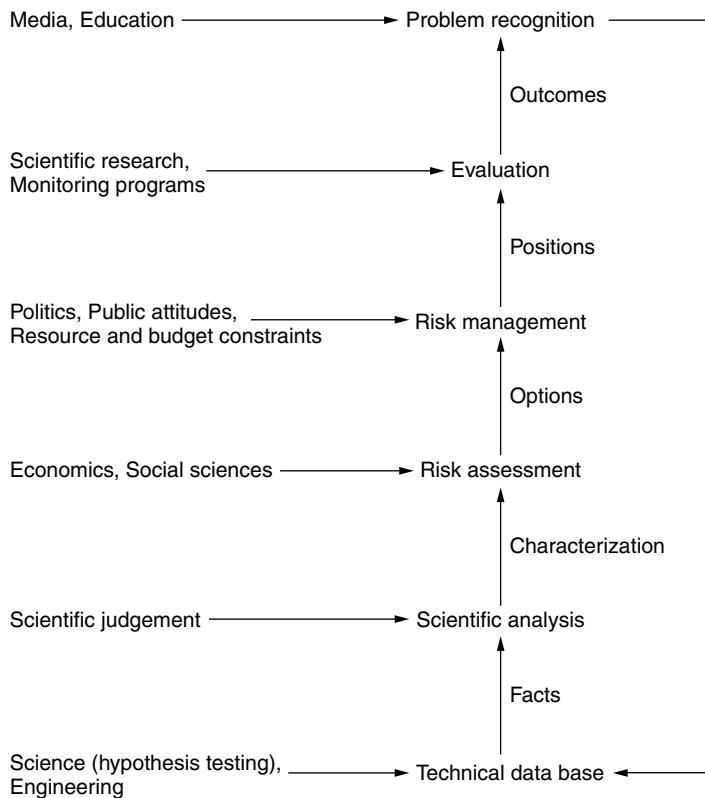
investigators at Stanford University in a study of dialysis and disability (as reported in *Time* magazine, May 20, 2008). A more common figure is about \$1.5 million, as calculated by investigators at Princeton University, in a study related to highway improvements.

There are also dimensions that do not lend themselves to scaling in these terms. Human suffering and environmental degradation cannot be easily translated into dollars (or euros, yuan, riyals, or pounds). Aesthetic enrichment also eludes easy conversion. As a consequence, the analysis tends to be driven by factors such as economic costs, which are more easily identified and estimated. The relatively intangible costs are more easily ignored, disputed, or subject to distortion in their interpretation.

Comparability is also a serious problem with current approaches to quantitative risk assessment. It is not possible to compare dollars against a human life in a way that is convincing to the general public, although this is regularly done by actuaries in insurance matters. It is also difficult to compare experiences that are qualitatively very different, such as dying of cancer compared to dying from a motor vehicle incident. Comparability is partly a problem of quantification, as noted above, but also involves social values and personal preferences that are at least in part culturally determined.

Science is where risk assessment begins but not where it ends. Scientific information is never the deciding factor in resolving occupational and environmental health issues. Rather, the scientific facts of the case provide a framework within which projections can be made and options for decisions can be explored. Making those decisions is fundamentally a social process, involving political, legal, economic, and cultural mechanisms. Risk assessment supports public policy by characterizing its consequences and the options available. Risk assessment can never replace the social consensus or decision-making process that a society has chosen to mediate among its various interests. It can only support those decisions.

Figure 7.6 provides a generalized schematic representation of the social management of problems involving risk analysis. The process begins at the top with recognition of the problem. This usually



**Figure 7.6.** Simplified risk assessment model.

occurs as a result of scientific research, hazard detection during environmental or occupational monitoring, education regarding an existing problem, or the attention of the media. Once identified, the problem becomes a topic for public discussion and scientific investigation. Public discussion may influence data gathering, especially if it is based on surveys or subjective opinions. The scientific task is often made much harder because the public inevitably tends to anticipate the results. Even so, a technical database is developed as objectively as possible, using the usual experimental designs of science and the technical approaches of engineering. This technical database is, of

course, argued over, digested, and reviewed within the scientific community. Issues of methodology and completeness may well be invoked to call some of the findings into question. Some consensus on the facts of the case usually emerges, however, and becomes the body of knowledge on which further analysis is based. Analysis is a more subjective process involving a good deal of scientific judgment to determine the significance of the findings and to integrate them into an overall concept of the problem. At this stage, experts in the field usually convene in meetings, workshops, and expert panels to discuss the issues and to construct what is, in effect, a framework for thinking about the problem.

## **RISK PERCEPTION**

The magnitude of risk in abstract terms means little to the average person. Studies consistently show that the average person tends to overestimate small risks (such as the risk of death from terrorism events or in an earthquake) and to underestimate higher risks (such as the risk of an injury from a crash on the highway or medical misadventure). This tendency is also present, to a lesser degree, among experts. Clearly, the perception of risk is not based on a rational assessment of probabilities. The field of risk perception is a systematic approach to how people process information regarding hazards they face and the implications for themselves, their loved ones, and their community. It is a broad field incorporating aspects of cognitive psychology and sociology.

Risk comes in two categories with respect to its psychological impact. This can be seen in the public's response to different types of disasters. "Natural" disasters, such as hurricanes, earthquakes, ice storms, mudslides, tsunamis, and volcano eruptions, are perceived as "acts of God"—beyond human agency and outside the control of the community, government, and certainly any individual. "Technological" disasters, on the other hand, are perceived as "acts of man," carrying a burden of moral liability on the part of the person, company, group, or agency responsible. Many comparative studies have demonstrated that people, across cultures, tend to adapt and recover

much faster from a natural disaster. This was made apparent in the recovery following the devastating earthquake in Sichuan in 2008. Neighboring communities, the country, and the world as a whole tend to respond with empathy and cooperation in the event of natural disasters. Technological disasters, on the other hand, are perceived (literally and figuratively) as “dirty,” involving messy implications of liability. They are prone to litigation, and victims often find their neighbors and the community acting ambivalently and sometimes with hostility toward them, especially if the perpetrator or responsible party is important to the community, perhaps as a major employer. Technological disasters are associated with blaming behavior, in part because of the legal liability and litigation they engender, which may seem perpetual. In addition, a natural disaster such as Hurricane Katrina can turn into a technological disaster, at least in terms of perception, if the public believes that the response or preparation was inadequate or poorly handled. This same pattern of perception is observed in relation to other hazards, but it is most obvious and best studied in the response to disaster.

Human beings are hard-wired to perceive threats and to respond to them in certain ways that are programmed for defense and protection. One aspect of this is the generalized stress response (see Chapter 13), but equally important are the responses of heightened vigilance and emotion. Prolonged heightened vigilance may be associated with anxiety in extreme cases but under normal circumstances will more likely be expressed as close examination and suspicion. Emotion is a fundamental part of the human response to threat and uncertainty, and it is normal to become angry or fearful when confronted with a threat. In addition, there are certain shortcuts in information processing in the human brain, called “heuristics” in psychology, that allow people to react quickly in an emergency but that bypass rational thought and information processing. One of these heuristics is the tendency to label or perceive new developments or objects in terms of dichotomies involving safety, such as good or bad, right or wrong, friend or enemy, safe or dangerous. As a consequence, environmental risk issues tend to carry many associ-

ations and connotations of morality and justice, and occupational risk issues tend to be interpreted in terms of workers' rights and fairness. This is entirely appropriate, but the tendency to force these issues into a moral framework can lead to a preoccupation with questions of "who is right" and "who is wrong" that neglects practical solutions and impedes progress in resolving the issue.

Against the backdrop of these heuristics, the perception of risk is modified by the processing of information within a learned context, a personal framework that is significant to the individual. A person with an advanced education in science will process information about probabilities very differently than will a person with little formal education. (Formal education is not the best marker of numeracy in any case. Scientists can have trouble balancing their checkbooks or reading a balance sheet; high school dropouts may become very skilled at handling money and complicated business transactions. Some of the best minds for probabilities are gamblers, who keep playing although they should know they cannot win in the long run.) That does not mean that the person with less formal education is not processing the information rationally. The person may be processing the information in a way that is rational given his or her level of understanding and experience. However, it is certainly easier to understand most environmental or occupational risk-related issues with the benefit of some education in science. This makes science literacy one of the major limiting factors in risk communication, as will be discussed later.

In addition to education, other learned attitudes and behaviors affect the perception of risk, at the level of the individual and of the group. Culture is chief among them, reflecting collective attitudes and group history. As a broad generalization, communities closer to their rural roots tend to be more accepting of risk. Communities that have experienced hardship recently may find small risks intolerable, perhaps because they have deep fears, usually expressed in terms of politics, about instability. This was apparent in the 1990s in Eastern Europe and in Russia after the collapse of the Soviet Union, when concern over environmental contamination became

a driving political issue, in the face of other, equally or more serious problems.

Whether a hazard is generally accepted or considered intolerable within a culture is also partly a function of demography and affluence. This has been the history of the United States and Canada in the twentieth century and is being repeated elsewhere today, most obviously in China. As societies age and become more affluent, they become more risk averse and intolerant of imposed risk. This is illustrated by the change in public attitudes toward impaired driving, which has occurred only within the last two decades. Whereas once the practice was widely tolerated as unfortunate but inevitable behavior, in recent years society has decided collectively that it will act punitively against the offender and has adopted numerous measures to curb the problem. In occupational health, attitudes toward work-related injuries and illnesses have also changed considerably. In economically depressed areas or during times of high unemployment, occupational injuries and illnesses are often accepted as a concomitant of employment. At other times, however, a worker's right to a safe workplace becomes the dominant focus, and previously sanguine members of the community may be outraged to discover a hazard in their midst. This tends to be more characteristic of newly identified hazards than of more familiar ones.

Once perceived, a risk may or may not be accepted. Many factors affect the acceptability of risk after the public, or individual, has rendered its verdict. Again, the expert tends to consider factors such as the benefits accruing from various action options, priorities for allocating resources, political considerations, public visibility, and the feasibility of control technology. These factors, while of interest, do not necessarily apply to the lay public and are not necessarily the basis for acceptability once the risk is understood.

Table 7.4 is a list of factors known to be associated with intolerance of risk. It summarizes many years of work by leaders in the field of risk perception, including Paul Slovic, Peter Sandman (who is most closely identified with the concept of "outrage" as a motivating factor in intolerance of risk), and many others. For the most

**Table 7.4.** Characteristics of Risk Issues Predictive of a Negative Perception, Intolerance, and Unacceptability by the Public

- 
- High impact (major consequence if the adverse event occurs)
  - High probability (high likelihood of the adverse event occurring)
  - Difficult to understand
  - Lack of familiarity (a novel or exotic risk, such as a new chemical or technology)
  - Unpredictability and uncertainty (the risk cannot be reliably anticipated)
  - Lack of control (especially by the individual, over personal or family's protection)
  - Involuntary assumption (people did not agree to it)
  - No benefit for those assuming the risk (if there is a benefit, it goes to someone else)
  - Acts on children (places children at risk now or in the future)
  - Acts on people who can be identified (people with names and faces, not statistical abstractions)
  - “Dreaded” or feared outcome (such as cancer, permanent disability, death)
  - Lack of trust (in the government or in the company responsible for the risk and its control)
  - Media attention (amount of general attention given to the issue as well as negative coverage)
  - Maldistribution of or inequity in risks and benefits (perceived as an environmental justice issue)
  - Irreversible effect (consequences will continue in the future and cannot be corrected)
  - Stake in outcome is personal (involves self or family)
  - Party responsible for imposing the risk can be identified
- 

part these factors are quite logical, and many are based on rapid processing through mental pathways (heuristics) that are hard-wired in the human brain rather than learned. It is perfectly normal, for example, to be more frightened of a novel hazard, such as a new chemical, than of something that is familiar, such as traffic hazards,

even if the risk associated with the novel hazard is actually substantially lower in magnitude. Its novelty means that the individual (and usually the community) has not had sufficient experience to evaluate the new risk and cannot know how to adapt, manage, mitigate, hedge, or otherwise accommodate it in their lives.

The pattern of the risk is critically important in determining public acceptance and shaping concern. A chronic, low-level risk is much more likely to be tolerated than the rare risk of a catastrophic event. The public is much more accepting of fossil fuel-powered electrical generating stations than they are of nuclear power stations, for example, although the former are associated with emissions contributing to air pollution and consumption of resources and the risk of climate change due to carbon dioxide emissions. The latter, at least those of modern design and operation, have an enviable safety record but are perceived to have the potential for unacceptable, catastrophic failure. (Actually, it is the problem of radioactive waste storage that is more limiting.) A rare event that carries the potential for a major impact is generally perceived by the public as a greater concern than a continuous or sporadic event of lesser consequence, although many more people may be affected or even killed in the latter case.

When there is uncertainty over the magnitude of a particular threat, the public tends to look at the uncertainty (at least presumptively) as a disturbing property of the problem they are worried about. Uncertainty is frightening because it implies unpredictability, which in turn implies that a risk cannot be reliably controlled or its realization anticipated. As a result, the public perception of risk tends to be unstable, changing with the week's headlines and the issues of the moment.

Voluntary or involuntary assumption refers to whether the risk is taken on by choice or imposed on a person, and it is a major determining factor in the acceptability of a particular risk. It is often said (although not well established) that the public will accept a 1,000-fold difference between those risks that are assumed voluntarily as part of one's lifestyle and those that are imposed by circumstances of employment or living in a particular community. Most people will-

ingly assume the risk of skiing or bicycling, both highly popular sports associated with significant (3 and 1 per 100,000 per year, respectively) risks of death but would protest job conditions so dangerous. A number of people are still highly reluctant to fly, and plane crashes are given considerable attention by the media despite the observation that the population risk of death from air travel is much less (0.3 per 100,000 per year) compared with motor vehicle travel on roads (22 per 100,000 per year); however, some critics point out that the better comparison is by number of trips, not time or distance covered. There is certainly no political consensus to ban hang gliding, although it is far more dangerous than any of the above-mentioned activities (80 deaths per 100,000 per year).

The public does not view cancer and heart disease alike, regardless of the dollar cost of medical treatment. The risk of an outcome that is dreaded, such as cancer, is considered much less acceptable than the same level of risk for a less dreaded event, such as a heart attack. In general, hazards that are very familiar, such as driving on the highway, are much more acceptable to people than new or exotic hazards, such as potentially toxic chemicals or nuclear power. Passive smoking is a very familiar hazard encountered by most people in North America repeatedly during their lives, but it took many years for this hazard to be separately identified and accepted by the public as a risk to nonsmokers.

Perhaps the most fundamental factor in risk perception, however, is trust. If people, as individuals or collectively as a community, do not trust the source of information, one or the other of the parties involved, or the one responsible for making a decision on what to do, no amount of objective fact will dispel their doubts. Trust is fundamental to the acceptability of risk.

This discussion has focused on environmental and occupational risks, which are almost always adverse. As noted earlier in this chapter, risk can also be perceived as positive, even as a commodity to be traded and taken on in anticipation of future profit. The insurance community is based on the commodification of risk. Likewise, financial institutions, until the housing securities–driven economic crisis of

2008, have traded in risk and hedged risks in the form of financial derivatives as a source of income and, for a time, wealth creation. To do so required trust at many levels—for valuation of securities, for trading, in the reliability of financial markets, and in the capacity of institutions to regulate their exposure to risk. While it is safe to say that this attitude toward commodified risk is certain to change, given the damage it has caused, a major cause of the crisis has clearly been the evaporation of trust. The experience of the financial sector demonstrates the centrality of trust. It is no less important, although perhaps less obvious, than other factors in the perception of risk by communities and individuals.

Risk perception has immediate practical application in the field of risk communication. Communication reflects the perception of the person sending the message; is understood, rejected, or distorted according to the perception of the person receiving the message; and is shaped by the perceptions of both parties with respect to the appropriateness of the medium of communication. Thus, risk perception is the foundation upon which risk communication is built.

## **RISK COMMUNICATION**

The general theory and practice of communicating information about risks and the management of risk-related issues is called risk communication. Although some observers consider it to be part of risk management, risk communication is also critical to risk assessment. It is fundamental to discussing and informing about risks at every level, whether between peer experts, expert and decision-maker, expert and the public directly, or decision-maker and the public, and in communicating via the media.

OEM physicians are frequently called upon to communicate on matters of risk. They represent their organizations, speak to workers, act as a resource at public meetings, give briefings to decision-makers, communicate information to their technical equivalents in other fields, and interact with peers. In all this, the OEM physician serves as the “expert,” the person with special knowledge but who

is not the decision-maker. This is a very precise role, involving special rules to learn and skills that are necessary to practice and master.

Effective risk management requires communication of the magnitude of the risk and the rationale behind either the steps required to bring the risk under control or the decision not to take any steps. It also requires communication of a framework for thinking about the risks and for putting them into perspective. The day is long gone when such communication could be one-way, directive, or didactic and still be effective. In contemporary society, the public expects and deserves to be part of a conversation, not the recipient of a lecture. This presents a special challenge to scientists, engineers, and physicians who are more accustomed to talking among their professional peers and using specialized jargon.

The public perception of risk is very different from that of experts. The difference lies not just in the vocabulary and mastery of technical details. Experts see the pathways; the public sees impenetrable complexity. Experts see exceptions; the public sees generalities. Experts see systems; the public sees one problem at a time. Experts see options; the public sees one “right” solution. At any time, the recollection of past impressions and a rumor of uncertain merit may negate the assurances of any authority figure. Lacking a basis for accepting an authority figure’s professional credentials, the public instead rely on their impressions of the authority’s credibility, a product of his or her media “presence” and apparent command of the vocabulary, concordance with one’s past knowledge or impressions, the reputation of the sponsoring agency, and numerous other factors that may be extraneous to the strength of the evidence at hand.

An expert will be held accountable for what he or she says in public, regardless of whether the expert is a public servant or a private citizen. Upon stepping up to the platform, the expert assumes ownership of the problem. The role of the expert is to help the public understand, to inform, and to provide frameworks and contexts within which to think about the problem. It is not to spin or to

**Table 7.5.** The “Seven Cardinal Rules” of Risk Communication Developed for the U.S. Environmental Protection Agency

- 
- Rule 1. Accept and involve the public as a legitimate partner.
  - Rule 2. Plan carefully and evaluate [one’s own or staff’s] performance.
  - Rule 3. Listen to your audience.
  - Rule 4. Be honest, frank, and open.
  - Rule 5. Coordinate and collaborate with other credible sources.
  - Rule 6. Meet the needs of the media.
  - Rule 7. Speak clearly and with compassion.
- 

Source: Covell V. *Seven Cardinal Rules of Risk Communication*. Washington, DC: Environmental Protection Agency; 1988.

manipulate the crowd or to push them to accept the decision already made by the expert’s employer. If the expert tries to do so, he or she will cease to be an expert in the view of the audience and will be seen as a hack.

### **The Basics: Seven Cardinal Rules**

The most common approach to effective risk communication is summarized in Table 7.5. The “Seven Cardinal Rules” were formulated in the 1970s for the U.S. Environmental Protection Agency by a pioneer in risk communication, Covello. They were developed in response to a conspicuous lack of success the agency experienced when it sent its experts, scientists, and engineers to public meetings to explain environmental issues. Covello has trained tens of thousands of people in government and the private sector in this approach.

The Seven Cardinal Rules are as follows:

1. *Accept and involve the public as a legitimate partner.* The expert is expected to demonstrate respect for the public, and failure to do so will generate a backlash, especially at public meetings, that at times can be hostile. It may seem odd that in the 1970s

experts talking about risk could still believe that the public should put blind trust in technocrats to solve problems for them. In a democracy people expect and demand to participate in decisions that affect their lives. They may, however, decline to do so. Sometimes that is because there are other, more pressing issues of concern to the community or because it is depending on activists or government agencies to keep the parties “honest.” That is their right and prerogative, and lack of engagement should never be mistaken for lack of interest.

2. *Plan carefully and evaluate your efforts.* One should never go to a meeting unprepared; it is better to find an excuse not to go than to arrive without a firm grasp of the issue and a clear idea of purpose. The public will hold the expert accountable for what he or she says and does. The expert should have objectives in mind in attending the meeting (key points to get across, the process that will be followed, recording issues that concern the community). Information on the issue should be reviewed and evaluated well ahead of time and the gaps, weaknesses, and uncertainties clearly identified. If at all possible, the expert should try out the message and delivery ahead of time on “real” people (not other experts) who are unfamiliar with the issue and who are as much like the expected audience as possible. One should never delegate responsibility for any presentation, including educational sessions, to someone who lacks the authority to answer questions. The final version of the presentation should always be the expert’s own, personalized and familiar.
3. *Listen to the audience’s specific concerns.* One should never assume what people know, think, or want. The easiest way to find out is to ask them, preferably by mingling with the audience before the meeting. People are usually concerned mostly with fairness, trust, credibility, competence, and control, not the technical details of an issue. They are also concerned that someone listen to them and take them seriously. Most people are profoundly uninterested in process. Timetables, approval deadlines,

decision points, schedules for public consultation, and legal authority sound like bureaucratic dodges to most people. The expert should keep these necessary elements of a presentation brief and straightforward, and concentrate on what is on people's minds.

4. *Be open, honest, and frank.* Trust and credibility are the most important assets of an expert; once they are lost, they are impossible to regain with the same audience. (Increasingly, communities and activists are comparing notes, so credibility lost before one audience may also have dire consequences later.) Many studies have been done on the credibility of sources of information and opinion, and the results are remarkably consistent. Physicians have the most credibility in the eyes of the public, sometimes even on matters where they normally have no special expertise, such as environmental issues that do not affect health. Scientists and university professors are also highly credible. Government agencies and nongovernmental organizations (NGOs) have less credibility, and distrust of public officials is spreading and deepening. Industry representatives and lawyers have the least credibility, and the OEM physician representing his or her employer should make every effort to act as a professional. To maintain an acceptable level of credibility, it is imperative that the expert not make a blunder in speculating, dissembling, guessing about a fact, or distorting information. If the expert does not know, he or she should say so and then promise to find and convey the information later, if possible. It is perfectly acceptable to say, "I do not know" when one does not, but it is never acceptable to guess or to mislead. Efforts to conceal a lack of information almost always result in tangles of dissembling, mistrust, and eventually hostility.
5. *Coordinate and collaborate with other credible sources.* When there are representatives of several parties at a public meeting, such as companies, government agencies, NGOs, unions, and other stakeholders, it is important to develop relationships with as many of them as possible. Cordial interaction and respect makes

communication much easier, even when the parties disagree. Experts with a common interest should agree on the facts and coordinate their message whenever possible, so that the public can focus on the decision or the interpretation of the risk, not on who is more likely to be credible. Sometimes this is not possible and there will be a public disagreement. When that happens, it is best to simply acknowledge it and not argue, to be respectful and to state the expert's position clearly while avoiding a public argument, and to let the other party explain why the inconsistency is worth the fuss. Unless there is a key fact or interpretation at issue, correcting the record over a detail should not be made into a big deal. When stakeholders who are more or less on the "same side" are not cooperating, it becomes even more important to monitor the media to see what the other parties are saying publicly. It is usually not the expert's role to direct strategy, but it helps to think through the issue from the other stakeholders' points of view. Are they holding back for a strategic reason or positioning themselves to break ranks? Is there another agenda? Is one agency trying to keep a low profile so that the other agency takes the heat?

6. *Meet the needs of the media.* The traditional print and electronic media are still generally the only effective means by which to communicate a risk message involving an environmental issue. (The public media are usually not very interested in occupational risks.) The media write the permanent record of an environmental risk and choose which of several stereotypical scripts will be used in describing the issue. For the message to get through accurately, the expert must view the media as a legitimate partner and not an adversary, even if the media have seemed biased or gotten the story wrong in the past. This means that the expert must consider the needs of the media for information and supply them in ways that will best ensure that the accurate and complete message gets through intact. The media have limited time or space and, for the most part, little capacity to sort out complexity. Journalists look for conflict, interest, and

a connection to the lives of their readers and viewers. They also need a “hook” (some compelling feature) to pitch their stories effectively to their editors.

7. *Speak clearly and with compassion.* Jargon has many important roles, among them the rapid and efficient communication of technical information among peers and the exclusion of people who do not belong in the conversation. Too many experts are unable to switch to normal language, however, when jargon is inappropriate. The ability to adapt communication styles is essential in effective risk communication. It is also helpful, whenever possible, to use concrete examples, vivid metaphors, and analogies that all people can understand. Most people respond to stories, not to statistics, facts, or theories. It is essential to respect and honor emotions, even if they are expressed in ways that the expert does not understand or finds inappropriate. The expert does not have to validate the emotions of a member of the audience who gets carried away or an audience expressing disbelief or hostility. It is critical, however, that the expert be seen as caring and empathic, not as detached and remote or retreating into a professional demeanor. A dismissive response by the expert to strong emotions in the audience will lead to hostility and outbursts of anger. However, the strongest negative reaction usually occurs when experts ignore people’s feelings and act aloof.

### Beyond the Seven Cardinal Rules

Although the Seven Cardinal Rules are still observed, they represent a very dated approach. The assumption behind them is that the expert knows best; his or her task is to convey complicated technical information to a public that is ill prepared to receive it. The conventional approach to risk communication assumes that there is an evident “truth” known to the expert and that the problem is one of persuasion. This approach is seen as patronizing by practitioners who believe that the role of risk communication is to provide the public

(or “publics,” since there are many groups and subgroups within the community) with the information they believe they need to make a decision, not solely what the expert thinks they need. The conventional approach treats culture as a barrier rather than acknowledging it as the organizing structure that people need to make sense of new information. It focuses on one item at a time, although people have many things on their minds and most communities have more than one environmental issue to deal with at the same time. The conventional approach is also becoming less effective, because the public (particularly activists) have been exposed to it so often that they know the moves and the script by heart.

Newer approaches have emerged from experience and study of communication theory and the cognitive psychology of risk communication. One of the most promising is the approach of “mental modeling,” which attempts to ascertain the landscape or vision of the world as perceived by the recipient of the message and what that person needs in order to act on and integrate the risk message. Risk communication is not a set of rules for “selling” a point of view on occupational or environmental risks, however. It is a framework for understanding the public’s view of such matters and putting forward an explanation that people will comprehend and evaluate on its merits rather than on relying on assumptions about the motives of the speaker.

People come to the table with prior experiences, beliefs, personal knowledge, culture, and all the wisdom and privations of life. They will look for meaning in what the expert says, how it is said, and how the expert is dressed. An expert who condescends to them, uses jargon, dresses in a manner seen in the community only on morticians, and acts as if his or her role is to educate them will be viewed as irrelevant. An expert who tries hard to fit in, uses speech that members of the community may use among themselves but is inappropriate for outsiders, dresses down for the occasion, and tries to play the role of advocate or partner (unless chosen by the community) will be viewed as a phony. Experts are best advised to speak naturally (as if to their child’s grade school teacher) and to explain

each technical term every few times it is used, to dress in keeping with their role (although scientists should leave their white coats in the lab), and to connect with the crowd by asking questions and showing genuine interest.

The alternative to the conventional approach is to treat the public as an equal to the expert. The expert has mastery of one domain of knowledge, but it is also important to learn what the listener needs to know. The audience's culture does not have to be a barrier to communication. If understood, it can be a "channel" through which communication can be made clearer and more efficient. The key is to focus on the life experience of the listener, not the level of education. Level of education is best regarded as a formality, bearing mostly on a person's fit within a particular social role and the associated opportunities available. It says little about one's ability to cope with complex problems and to make decisions. The more that can be learned about the culture and point of view of each of the "publics" that make up the audience, the more easily a message can be conveyed in a way that it will be informative and understood. At the same time, the expert should beware of the tendency toward cultural stereotyping. The best way to avoid this is to get to know individuals in the audience and to learn from them.

### **Practical Skills for Public and Media Appearances**

Most educated people are adequately equipped to form general opinions on the principles behind OEM issues but lack the technical knowledge to unravel the issues in their complexity. Instead of weighing evidence according to its scientific merit, most people are moved to accept or to reject a proposition on the basis of the perceived sincerity of the speaker, the assumed motives of the institution the speaker represents, the plausibility of the argument, the degree to which it conforms to their expectations, and the congruence of the conclusion with their preconceived notion of what action should be taken. The essence of risk communication is the same as for any other form of communication: to communicate knowledge, understand-

ing, and compassion. The expert has to work hard to ensure that the information is intelligible, that the framework for understanding is established, and not only that the audience is reassured that the expert cares about them, but that the expert actually does.

The expert should think through the essence of the environmental risk message, reduce it to its fundamental elements and to the degree of simplicity that the material allows, and prepare a few key talking points. The expert should be accessible to the media and, when asked for an interview, should speak to these key points in simple and direct language and short sentences. The ability to speak in concise “sound bites” is a legitimate and necessary communication skill that can be practiced and mastered. Media “kits” should be distributed whenever possible, containing all the information a reporter needs to write the story, including background information and technical explanations. Eventually, as journalists become comfortable with the expert and come to trust him or her, the expert will be called upon earlier and will be more effective.

Media training is an essential investment for anyone engaged in risk communication. The traditional media are now supplemented with Internet Web sites, blogs, and podcasts, among other formats. Determining how to use these new communication channels effectively is a major challenge in risk communication.

Body language is particularly important. Some communication experts believe that 90 percent of communication is nonverbal. The public tends to read messages into how one sits and behaves during a meeting. It is important to avoid facial expressions that may reflect an opinion (scowling or grinning), gestures of distraction or impatience (such as drumming fingers or tapping feet), signals of inattention (such as resting an elbow on the table to support one’s head), suggestions of defensiveness (folding one’s arms across the chest or leaning backward), and appearing socially awkward (swaying or rocking back and forth at the podium).

Many of the issues involved in occupational and environmental regulation are dense in detail and complexity. The material should be made as understandable as possible to the intended audience but not

simplistic or condescending. Know the audience, including its general level of education, language skills, and experience with the topic, and adjust the presentation accordingly.

A particular problem in this field is that many expressions of risk are couched in probability statements or expressions of uncertainty that are difficult to translate into simple terms. To a layperson, the use by experts of terms such as “insufficient evidence,” “low probability,” “all other things being equal,” “nonsignificant,” and “non-causal association,” in addition to all the qualifiers attached to technical statements, sound like equivocations. In speaking to the public and to workers, it is important to lay out what is known and what is not known in simple language, even if it takes more words, and to avoid piling on what the public perceives as “weasel words.” Rather than “evidence suggests that the probability is low that there may be an effect which would in any case occur only under certain circumstances,” the expert should say, “what we know now makes us pretty sure that nothing will happen in general, but we’re not sure about one particular situation.”

In any public forum, it is critically important to develop a rapport with the audience. This is necessary in order to establish a relationship in which members of the audience perceive that they are being spoken to as individuals rather than as a mass to be soothed or a constituency to be guided. The speaker or writer who fails to impress the public with his or her personal sincerity as well as scientific authenticity will not receive a hearing.

The speaker should show respect for the audience at all times, even when under verbal attack. To do otherwise is to invite the perception of arrogance and hostility. The representative sent to a meeting should be appropriate to the occasion. For example, one would not send a mid-level manager to testify at a congressional hearing. Likewise, one should not send the most junior person on the team or a public relations representative to meet with community leaders about an environmental problem or deal with union representatives through lawyers.

Questions from the audience should always be encouraged, and ample time for them should always be allotted at the end of the pro-

gram. Interruptions during the program itself should be discouraged to allow coverage of all the necessary material, but the audience should not have to wait too long before being given the opportunity to express their concerns. Many people will have attended because of a pressing personal interest and will have opinions, attitudes, or concerns that they wish to express. They may feel squelched and frustrated if they have no opportunity to express themselves.

This also provides an opportunity to find out what the audience thinks of the issue. Only by understanding the depth and intensity of their opinions and motivation can one begin to comprehend the significance of the issue and the degree of approval or resistance a proposed action is likely to face. The public may also be aware of alternatives or may be able to provide very useful practical information that may determine the success or failure of a proposed option. Ultimately, it is the community, rather than the expert, who will have to live with the consequences of the decisions made. It is best that they have input into the system and are in agreement. At the very least, the dissenting minority must feel that it had its day in court so that the resolution is not destined to many days of litigation in court.

Although measures to establish a rapport with the audience will increase the chance that the message will receive a hearing, they do not guarantee acceptance of the proposition under discussion or the point of view of the speaker. Effective risk communication does not mean that the public ultimately agrees with the assessment of risk put forward or with the means proposed to control it, only that the reasons behind the proposal are understood. In a democratic society, the people always retain the right to disagree, and using that right should never be dismissed as an expression of ignorance.

Whatever statements about risk are put forward, there will almost always be dissenting views. Most people perceive dissent as conflict and interpret lack of agreement among scientists, in particular, as confusion and uncertainty. Lacking the means to weigh the sides of the argument, the knowledge to weigh the credentials of the authorities involved, and the time to focus one issue in an otherwise busy life, most people look for simplifying assumptions. These may take

the form of a conclusion that it is better to be safe than sorry, a conviction that the issue is not so important after all, a desire to let government regulation handle the matter, or a wish to sidestep the issue altogether by going in a different direction.

### **Literacy Levels in Written Communication**

Language can be as much a barrier as an avenue to communication. Experts tend to write about environmental risk messages in their professional jargon or, if they are sensitive to the need to use accessible language, to lapse into specialized vocabulary when they get stuck. Physicians are no different from engineers in this tendency. Complexity of syntax and vocabulary defeats good communication with the public, however, and must be resisted. The literacy or readability level of the material must be appropriate for the readership.

Readability levels are not exclusively about education. Although persons with more education are more likely to understand complicated language, even those with a high educational level may have trouble if they are stressed, frightened, tired, aged, or ill. People whose first language is not English are even more likely to have trouble with comprehension when under stress.

Material written for the public can be evaluated for reading level with the use of a simple, non-proprietary test called the Fog Index. It is described in Table 7.6.

Material designed for the public should be rich with graphic illustrations and “white space,” employing wide margins around the text. The text should be in an easily read, large font (especially if it will be read by older people), reproduced in black on white paper for sufficient contrast. Documents should be as short as possible—preferably one page, front and back if necessary.

### **Worker Education and Right to Know**

The OEM physician is often called upon to communicate information about job-related risks to groups of workers. This may take

**Table 7.6.** Calculating the Fog Index

- 
1. Select one paragraph (more if the paragraphs are short) of more than 100 words. This should be the most difficult passage if the objective is to determine whether the entire document is readable.
  2. Count the words up to the end of the sentence that comes closest to 100 words into the paragraph. For example, if the paragraph has 630 words and 26 sentences, and the last word in the fifth sentence is word 95 and the last word in the sixth sentence is word 106, pick the end of the fifth sentence because it is closer to 100. (This procedure is designed to make percentage calculation easier.) This will be the sample.
  3. Count the number of sentences in the sample. This number is  $N$ .
  4. Calculate the average length of the sentences in the sample by word count. This number is  $L$ .
  5. Count the number of “long” words, those with three or more syllables, in the sample of 100 words. Do not count proper nouns (such as “Alpharetta”), compound words that are composed of shorter words (such as “decision-maker”), or verbs that have one-syllable tense-related endings (such as “disseminated” or “accelerating”). This number is  $W$ ; it is the percentage of “long” words in the sample.
  6. To derive the Fog Index, add the average sentence length and the percentage of long words and multiply by 0.4:

$$\text{Fog Index} = (L + W) \times 0.4$$

7. The Fog Index is equal to the level of schooling required of a person for him or her to read the sample easily. For example, a document with a Fog Index of 11 should be easily readable by someone who graduated or at least finished most of high school. Text with a Fog Index of 17 could be hard for a university graduate to read.

Material intended for the general American public should be written to produce a Fog Index of 7, corresponding to middle school, grade 7. Any material with a Fog Index higher than 11 is not suitable for the general public. Comic books have a Fog Index of 6, *Time* magazine 11, *Atlantic Monthly* 12.

The term “Fog” refers to clarity, not to the name of the person who created it. The Fog Index was developed in 1952 by Robert Gunning for teaching business communication. There are several variations, of which this version is the simplest.

---

the form of training or safety meetings (which do not usually involve physicians) or special meetings arranged to address recent problems or in the aftermath of an event.

“Worker right to know” is the concept that individual workers have a right to understand the risks to health that may accompany exposure to chemicals or physical hazards on the job. It presumes that the employer is responsible for providing such information to the worker as needed or requested. The process of providing this information is called “hazard communication” as a term of art, but fundamentally it is applied risk communication. OSHA has promulgated a Hazard Communication Standard covering training in chemical hazards, as well as labeling and material safety data sheets, which are discussed in more detail later. Hazard communication specific to hazards, occupations, and industries is also incorporated into several other OSHA standards.

Worker education or right to know assumes that employers know what hazards exist and the risks they entail; this places a duty on employers to find this out. It puts workers on a more equal standing with management by giving them information that can be used to challenge decisions or work policies and, if necessary, to justify refusal to work under unsafe conditions. Worker right to know presumes that someone, as a practical matter the employer, must take steps to ensure that workers are able to receive, interpret, and act on the information. Because of these ramifications, adoption of worker-right-to-know legislation has had a major impact on occupational health and safety regulation, often in indirect ways. In the absence of effective regulation, for example, it has allowed workers to come to grips with their own workplace issues and insist on their rights.

Recognition of an implicit right of access to information on the hazards that workers may encounter on the job was slow to come in North America but is now firmly entrenched. Workers may have a right to information on hazards in the workplace, but there must be some provision for interpreting this information in terms that a non-professional can understand. The principles of risk communication apply as much here as in public meetings on environmental issues.

The worker must be educated to understand the implications of the information. A hazard should be immediately recognizable without access to lengthy documentation. There should be some way of confirming that amelioration measures have been taken and some penalty if they are not.

When the Occupational Safety and Health Administration introduced the Hazard Communication Standard in 1987, it was very controversial and bitterly opposed. The standard was finally upheld after three years of litigation. OSHA enforced the standard quite strictly, citing and fining employers who failed to comply in detail. Almost simultaneously, the federal government and individual provincial governments of Canada introduced the Worksite Hazardous Materials Information System, a somewhat more detailed but generally comparable piece of legislation. They were the first two systems to be developed and were eventually harmonized for compatibility between the two countries. The European Union has since developed its own approach, and efforts are under way to develop an international hazard communication system, directed primarily at labeling and documentation.

The general outlines of hazard communication are similar across North America. Employers are required to maintain information sheets, known as “material safety data sheets” (MSDSs), on every chemical used or produced at the worksite. The MSDS for any chemical in use on-site must be produced on request by a worker. The basic information to be included in the MSDS is specified, although these documents are generally less complete and authoritative on the health effects of exposure than on explosion or combustion hazards. The employer is required to provide formal training for all employees who may be exposed to hazardous chemicals. Strict labeling requirements apply to hazardous chemicals, and a set of standardized, easily recognized symbols have been created to warn of particular types of hazards.

A side issue to the development of hazard communication systems has been the discovery of widespread illiteracy in the workplace across North America. It has become obvious to employers that

many native-born workers cannot read written materials with sufficient comprehension to comply with training requirements. Among some major employers this has led to crash programs to improve reading skills. A related problem is literacy in English for immigrant workers. Major employers generally provide training in Spanish and the other languages commonly spoken among their immigrant workers, but smaller employers may not.

### **Worker Notification**

Worker notification is a concept similar to worker right to know but applies following exposure to hazard in the workplace. Workers who have held jobs in industries known to present a relatively high risk of certain health problems, most obviously cancer, should be individually notified that they are at risk and should be advised to take appropriate action, such as undergoing periodic health surveillance, receiving early medical treatment for conditions as they develop, or modifying their risk through some intervention. (Asbestos workers, for example, might be advised to stop smoking if they do smoke and to continue the OSHA-mandated surveillance procedure for the rest of their lives, whether they continue to work in the industry or not.) However, there are many complications.

Responsibility for notification generally falls on employers, along with an implied duty to transfer employees to less hazardous work without loss of benefits. Unfortunately, this is often not done, or it may be resisted by the worker because the job requiring handling of the hazard may pay more. It has been proposed that employers be responsible for the costs of lifelong surveillance; this argument grew out of the need for employers to support surveillance to detect early bladder cancer in workers exposed to carcinogens in the chemical industry. This commitment to “medical monitoring” is often resisted by employers and may not be cost-effective except in situations of extreme risk, given the absolute lifetime risk of most occupational hazards compared to other sources of risk for the same outcomes. The case for medical monitoring has migrated to the courts, espe-

cially in cases of community-wide environmental exposure, and is often part of settlement negotiations in tort litigation.

There has been concern that workers and their families would be alarmed and psychologically stressed by notification that their risk is elevated, even when the magnitude of the risk is not great in absolute terms compared to other risks in life. This phenomenon, called “labeling” in the health behavior literature, has been demonstrated in patients presumed to be hypertensive, but it is not known how large numbers of otherwise healthy workers would react collectively and individually to being told that they might develop a health problem many years into the future.

Some countries have legislation requiring employers to inform workers if they are at high risk of exposure or have been significantly exposed to occupational hazards. In 1987 there was a bill to introduce such legislation in the United States, but it was strongly opposed by industry and was unsuccessful. Even so, some employers provide such information as a matter of responsibility.

Like worker-right-to-know legislation, the policy of worker notification places the worker on a more equal footing with management. With a worker notification policy, workers have the option to pursue litigation in the short term against employers on the basis of anticipated health risks, loss of insurability, or mental distress. Such a policy also gives workers the option to pursue litigation against former employers for health problems that are only equivocally occupational or for failure to notify. Notification provides the worker with an incentive and with the information needed to take steps to obtain medical attention and, if necessary, to prepare a legal case for compensation. Litigation is likely to be dominated by the presumption of risk, which would tend to give any worker in an affected job classification the benefit of the doubt, even if the exposure were brief and personal risk factors, such as smoking, were present.

Several questions remain to be resolved before acceptable worker notification procedures can be developed. One is how conflicting evidence on the presence and magnitude of excessive risk will be

handled in selecting those occupations to be designated as high risk. Another is whether notification should be restricted to workers who have sustained exposures resulting in outcomes that physicians can do something about. Is it, in other words, worthwhile to worry workers who may have sustained exposures in the past if there is nothing that can be done now to change the outcome? If not, who will decide when clinical intervention is sufficiently effective to warrant notification?

An example of a promising approach to early intervention, which ultimately proved disappointing, is chemoprevention for lung cancer and asbestosis. A clinical trial using retinoic acid, which had been shown in other studies to inhibit the development of epithelial cancers, was initiated in a large cohort of asbestos-exposed workers in Seattle. The early results showed that the intervention appeared to increase, not decrease, the number of detectable cancers, perhaps by accelerating the growth of those already present. The investigators immediately stopped the study and halted the intervention, of course, but the experience raised many of these same questions.

Another issue is whether notification should be based on actual personal or environmental exposure data, which is usually incomplete, or on job category. In the former case, many workers who were exposed will be missed because of incomplete records. In the latter, many workers will be notified who are not at particularly elevated risk. Workers in different job categories who happened to work in areas where sporadic exposure might have occurred might still be missed.

At present there are few examples of early interventions that would be effective for occupational cancers; survival from lung cancer, for example, does not seem to be improved by early detection, even at the current state of the art (spiral high-resolution CT scanning). Screening tests are also limited in their ability to detect the earliest, reversible stages of a disorder.

In the future, other modes of intervention may increase the number of occupational diseases for which early detection is worthwhile. In addition, newer screening modalities will make earlier detection

possible for some cancers and other chronic conditions for which early diagnosis is useful.

### **Community Right to Know**

In the 1980s several states and larger municipalities in the United States (including San Diego, California) adopted legislation creating a right on the part of citizens to access information on what hazards may be present in their local communities. For example, a city or county public health agency may be charged with the responsibility to procure mandatory reports from local industries on what toxic chemicals are used on site. Fire and public safety departments now inventory toxic substances in their communities, sometimes as part of community right-to-know legislation. Usually, there is a cutoff for reporting requirements such that small amounts of chemicals of relatively low toxicity are exempted. Local citizens then have the right to inquire and receive information on what hazards may exist in their neighborhoods and what steps might be taken in an emergency to mitigate risk. Proprietary information and trade secrets are respected, but provisions are sometimes made for overriding this protection in the event of an emergency such as a spill or fire. There are now a plethora of such local acts.

Community right to know has been greatly expanded by federal legislation, especially the Emergency Planning and Community Right-to-Know Act of 1986 (EPCRA). EPCRA was passed following the poisoning incident in Bhopal, India, and was expanded in 1997. It is mainly about emergency planning and community response plans. An annual Public Data Release allows tracking of risk issues in any community. Likewise, the Superfund Amendments and Reauthorization Act of 1986 (SARA), which mostly concerned updating Superfund, brought several more community right-to-know provisions, requiring disclosure of proprietary information to medical practitioners in cases of exposure and greater involvement of citizens in steering cleanup of hazardous waste sites. These and other community right-to-know acts now are coming into conflict with homeland security legislation and policies, since the information they cover can be sensitive.

Several EPA programs have expanded access to data on community risks. The EPA operates the Environmental Justice Geographic Assessment Tool, an online database that locates facilities, local institutions such as schools, and sources of pollution on maps. The National Environmental Monitoring Initiative provides information on potential (but not necessarily actual) exposure.

The chemical industry has taken a proactive approach with respect to community right to know. The Community Awareness and Emergency Response (CAER) program is an industry-wide effort to respond to community concerns before problems arise. It consists of public information programs, annual plant open houses, design and public announcement of disaster response procedures and contingency plans, and collection of feedback from the public on preferences and attitudes. Obviously, these programs are not substitutes for effective control of on-site hazards but can be important in facilitating cooperation in the event of an untoward incident. They implicitly recognize the community's right to know and involve the company in an ongoing process that may result in greater sensitivity to public perceptions.

## **RISK MANAGEMENT**

Once the findings of risk assessment exercises are transmitted to an appropriate decision-making body, the process of risk management begins. Risk management is properly the function of regulatory agencies and the parties responsible for the risk. Risk management depends on the legal authority of the regulatory agency, the liability or accountability of the party held responsible, and the technical options that are open for replacing the exposure or controlling it. This is a political, economic, and management decision, not a scientific problem to solve.

### **Deciding to Decide**

Occupational hazards tend to be visible and make their presence known sooner or later. The effects of environmental hazards are usu-

ally not obvious. Except for unusual and catastrophic events, exposure to chemicals in the environment acts more subtly and is rarely easy to identify as a significant factor in the health status of an individual. Because of this uncertainty, some standard of judgment is needed in order to decide when to act: What level of certainty that there will be health effects should be sufficient to trigger regulation and hazard protection, which, after all, comes with costs? At least seven criteria for action are now in common use:

1. *The precautionary principle.* This criterion holds that if there is a possibility of harm, there must be action to prevent a risk, whether there is hard evidence or not. A corollary of the precautionary principle is that protective action should be taken before scientific proof is possible in order to prevent unanticipated, serious consequences. The precautionary principle was introduced specifically into environmental affairs as part of the 1992 Rio Declaration and was popularized by a famous meeting (the Wingspread Conference on the Precautionary Principle) in 1992. It has become hugely influential, particularly in the European Union. However, because social, political, economic, and legal institutions are primarily moved by evidence and caution is difficult to enforce in the face of competing social benefits and economic gain, the precautionary principle is most often used for smaller scale risks, such as control of a carcinogen, rather than the large-scale, destabilizing risks, such as climate change, for which it was originally intended.
2. *The de minimis risk standard ( $10^{-6}$ ).* The EPA uses a probability of one cancer among one million persons over a lifetime as the standard for what is acceptable; given the usual uncertainties, it follows that a risk of 1-in-100,000 is the criterion for action. This criterion is arbitrary but seems to reflect a generally held social consensus that cancer, in particular, is a dreaded health outcome to be avoided even at high cost but that one-in-a-million is simply too remote a probability to worry about.

3. *The public health standard ( $10^{-4}$ – $10^{-2}$ )*. Public health authorities tend to use a poorly defined but higher cutoff for the acceptability of preventive measures, even when the risk of not using them is higher. This is illustrated by the risk of complications due to immunization compared to the risk of unimmunized children who are exposed to, for example, measles developing serious complications, such as pneumonia or encephalitis. Other examples include the risks associated with smallpox immunization or the risk of Gullain-Barré syndrome or allergic reactions to more common vaccines. This vaguer criterion also applies to the risk of inducing cancer through excessive radiation exposure in screening programs based on mammography.
4. *Clinical standard ( $10^{-2}$ – $10^{-1}$ )*. The implicit, usually unstated, criterion for accepting a risk of an adverse drug effect or a complication of treatment depends on the severity and prognosis of the illness, the availability of effective treatment, and the severity of the complication. This standard has not been extensively documented but underlies numerous decisions made every day in clinical medicine.
5. *Legal standard (>0.5)*. In law, the operative criterion in civil cases is the balance of probabilities in terms of “more likely than not,” a criterion that implies greater than 50 percent probability of a particular outcome being associated with a particular cause. To make a decision based on these criteria, one must be persuaded of the evidence for or against, but only just, on the “weight of evidence.”
6. *Scientific standard (>0.95)*. In a particular experiment or study, conventional significance testing assumes a possibility of error of less than one case in twenty. Of course, any chain of scientific evidence is based on many individual studies, most of which are likely to be replicated in whole or in part. Thus, the probability of the entire chain of evidence being in error is actually much less ( $<0.05^n$ , where  $n$  = number of replicate studies). This still allows for error but provides confidence in the data underpinning the conclusion. The scientific standard of

evidence is not feasible for most purposes, however. It is simply too stringent, and it takes too long to achieve such certainty. Decisions have to be made in real time with the evidence available, and a scientific standard of certainty is not usually possible.

7. *Beyond reasonable doubt.* This is the standard used in court in criminal cases. This is a very high standard, rarely attainable in other human affairs.

### **Risk Management Options**

An appropriate risk management decision takes into consideration the likely consequences of the various options presented and balances each, taking into account other issues of public concern. The political system of a society is usually the arena in which these trade-offs are made, although sometimes the legal system decides the issue in a prominent court case. In making this decision, politicians and other decision-makers take into consideration public attitudes and willingness to cooperate with changes in their life, constraints on resources and budgets, and competing interests of economics. Ultimately, a position will be taken, even if that position is not to do anything. It is axiomatic in regulatory policy that “no decision” is a decision. A decision not to decide is tantamount to allowing the status quo to continue or leaving it to market forces to resolve the problem.

Table 7.7 presents general approaches available for the management of risks that a decision-maker, such as a legislator or an official of a regulatory agency, might consider for a particular problem. They are the basic choices available for the control of hazards in the workplace or in the environment. (This formulation is adapted from a framework first advanced by the economist and risk scientist Lester Lave.)

Standards and guidelines are partial measures that allow limited exposure at levels presumed to be safe. The standard is, or should be, based on rigorous scientific research when it is available, but standards are usually modified by social and economic considerations. They may be set on the basis of the greater good to society (usually construed as economic), and this is formalized when standards are subject to a cost-benefit or risk-risk

**Table 7.7.** General Approaches to Risk Management

- 
- Market regulation, “freedom to choose”
  - No risk permissible, such as bans and absolute prohibitions (e.g., unsafe toys, Delany clause)
  - Technologically defined solutions, usually best available control technology (BACT) which refers to a requirement to use the most effective technology currently available for emissions control, taking cost into consideration
  - Risk-risk analysis, to choose the lesser evil
  - Risk-benefit analysis, to choose the greatest proportionate good
  - Cost-effectiveness analysis, to obtain the biggest bang for the buck
  - As low as reasonably achievable (ALARA), a requirement to exert every reasonable effort in design, technology, management, and operation to keep emissions as low as possible, ideally far below limits or standards
  - Stringent control to the maximum extent of technology, as low as feasible (ALAF)
  - Standards (mandatory) and guidelines (voluntary)
  - Continuous improvement (including nondegradation policies)
  - “As low as reasonably practicable” (UK), implying investment in control proportionate to magnitude of consequences
- 

analysis. Standards and guidelines, because of their great importance in OEM, are discussed in greater detail later in the next section.

“Best available control technology” (BACT) is the usual policy approach to high-level risks, such as ambient air pollution or nuclear energy.

“As low as reasonably achievable” (ALARA, a term that came out of health physics and radiation protection) is the norm for hazards that analysis deems to be relatively limited. In practice, occupational hazard regulation is usually much less stringent than environmental regulation because of the perceived risk-benefit balance, which assumes that healthy workers are less likely to be adversely affected.

The United Kingdom’s Health and Safety Executive (HSE) uses two additional terms: “as low as reasonably practical” (ALARP),

which is a more technical term implying a quantitative analysis, and “so far as is reasonably practicable” (SFAIRP), which is used in public guidance. Both mean that the cost of controlling a hazard should be weighed against the consequences and that proportionate response should be made, with increasing stringency and investment to reduce more serious risks.

Market approaches have been the norm for the control of exposure to tobacco-related and other lifestyle-associated hazards. These approaches depend on education and notification on packages to influence consumers in the individual decisions they may make. Market regulation respects the freedom to choose on the part of the consumer. Under this approach, the consumer is empowered with the choice, and government typically restricts its role to public information that educates the consumer about the extent and severity of the hazard. Once this educational function is discharged, the government is presumed to have no further interest in the behavior of the individual, and no regulation is applied. In the United States and Canada, market regulation is used implicitly in the consumer selection of a number of products that may or may not have a significant impact on the environment and is part of the framework for making choices about transportation and other issues of daily life. The area of consumer health in which the approach of market regulation has the most consistent application, however, is in control of tobacco products.

Some important applications of the market principle are the concepts of “cap and trade,” which is a pollution tax or fee paid polluting industries in proportion to the amount of discharge. This application is often called “polluter pays.” In theory, this creates an incentive to polluters to reduce discharges and returns to society the costs for cleaning up the environment. The idea has had little application in occupational health, as it would be unacceptable to allow an employer to expose workers to the payment of a fee. The principal problem in applying this concept is in knowing how to cost out the fees on a national basis. The “cap and trade” system forces polluters to buy permits (the number of which are limited

by a cap that is steadily reduced) to pollute, which they then trade among themselves in an open market. This system has the great advantage of maximizing technological and economic efficiency, as polluters use the market to seek out the most efficient ways to make short-term gains.

Outright bans are highly unusual in regulatory policy, as demonstrated by the resistance to banning asbestos in the present day, even when its use has become almost negligible. The exception is consumer products, which are banned most often for safety reasons.

An approach that historically has not worked very well is that of permitting no risk across a broad category of outcomes or class of hazard. This approach does not allow for discretion in balancing risks against benefits and assumes that the outcomes are such that no risks are acceptable. An example of this philosophy in practice was the Delaney clause of the U.S. Food and Drug Administration, a provision of the Food and Drug Act that requires that the agency prohibit the use of any food additive demonstrated to be carcinogenic. This provision does not allow the government to differentiate between substances and became untenable as more chemicals were found to have some activity in some animal studies, whether germane to humans or not, and as smaller concentrates became detectable with advances in monitoring technology. The Delaney clause is now applied more flexibly in cases of *de minimis* risk through a variety of administrative interpretations, and pesticide residues have been removed from its authority.

## Options Assessment

Another approach to risk management is to make a decision among several options based on an analysis of the relative costs and the relative merits of each. These analyses are the direct product of the risk assessment exercise and result in a series of well-defined options among which decision-makers may choose the most socially acceptable or the one that produces the optimum benefit. Such deci-

sions are not, by their nature, scientific but involve in large measure a set of intangibles related to those factors that are so difficult to quantify in risk assessment. These analyses are of four types: risk-risk, risk-benefit, cost-effectiveness, and benefit-cost.

In risk-risk analysis, the risks of various options are compared, and decision-makers choose that which is perceived as the “lesser evil.” For example, the risk of a certain chemical may be perceived as significantly lower than that of another, and the decision may be made to substitute the chemical presenting the lower risk.

Risk-benefit analysis evaluates the risk against the demonstrable benefit of each option in an attempt to identify the “greatest good” for society as a whole. For example, pesticides that are known to cause some degree of ecological disruption and to be acutely toxic are permitted in the marketplace because the benefits to the economy and to quality of life appear to outweigh their potentially small and well-defined risks. This type of analysis suffers from the general problem that risks are much easier to quantify than benefits.

Cost-effectiveness analysis is usually performed on a departmental or agency level to decide on the allocation of funds for the control of several hazards. If an agency such as OSHA is deciding what category of exposure it will concentrate on, it may determine what it would cost to establish standards for certain categories of hazards and may then choose an option that will allow it to get the most effective regulation for a limited amount invested. An example of this approach would be the decision by an agency to develop a generic standard for carcinogens rather than establishing specific standards for a variety of carcinogenic chemicals or for another class of toxic agents entirely.

Benefit-cost analysis is basically a matter of calculating a return on investment and is summarized by the phrase “as low as feasible” (ALAF). In this approach, funds are expended up to a predetermined or practical limit, but total control or an unlimited effort is not considered warranted. An example of this approach might be the establishment of an exposure standard for benzene or another chemical at

a level that might predictably result in at least some toxic effects. The resources to implement total control would be perceived as excessive, and the benefits of regulation would accrue mostly by a more modest expenditure.

An approach to risk management that has historical precedent but never seems to be explicitly adopted is “continuous improvement.” The history of risk management, especially with respect to standards setting for chemical exposures, is generally one of increasing stringency, reducing allowable exposure over time as new information becomes available on hazards. One wonders why this approach cannot be adopted as a permanent policy, such that at every opportunity technological and operational changes will be made that favor a reduction in emissions or exposure to hazard. At present, standards are set for everyone, on a rigid timetable, in order to achieve a level playing field. This has the effect of driving up costs and is inevitably opposed by the regulated sector, which imposes delays and may drive costs up further. Progressive improvement might come more easily if made at natural points in the cycle of business, for example, when a new plant is built, there is turnover in a fleet of cars, maintenance is scheduled on a plant, or when equipment comes due for replacement. A policy of continuous improvement, in which pollution standards are never allowed to get worse but may become more stringent, is applied in the EPA’s non-degradation policy for air pollution. Something like it is implied in the EPA’s ambient air quality standards review cycle. Otherwise, it is rarely put into practice. The reasons may be that in order to be sustainable, such a policy would have to be consistently followed for many years, through many successive administrations or governments, and with a strong degree of cooperation and compliance from the regulated sector. At the local level, timing- and opportunity-driven decisions would need to be made that only local operators and owners could manage. One may conclude that effective implementation would require too much social and political collaboration to be realistic.

**Table 7.8.** Criteria for Analyzing Risk Management Options

- 
1. Economic efficiency: Are resources available and can the job be done with what is available?
  2. Equity: Are those bearing the risk those obtaining benefit?
  3. Administrative simplicity
  4. Priority among other needs in allocating resources
  5. Communication and ease of explanation to public and decision-makers
  6. Scientific validity and technical utility
  7. Political acceptability (to legislators, voters, agency officials, residents, and all other stakeholders)
  8. Likelihood of success in a legal challenge
  9. Legislative mandate and legal authority of agency involved
  10. Incremental action (Could this approach lead to a better future alternative not now politically acceptable?)
  11. Continuous improvement (Is this action going to improve the situation in the future and prevent reversal?)
- 

Table 7.8 summarizes some of the considerations that a decision-maker might take into account when selecting a risk management option.

## STANDARDS AND GUIDELINES

Standards, in regulation and risk management, are binding and enforceable requirements, usually in the form of maximum allowance for exposure, that are imposed by a regulatory agency or other authority on a universal basis. Compliance with standards is mandatory. Examples of standards include OSHA standards in occupational health, EPA standards such as the National Ambient Air Quality Standards and emissions regulations, local and state discharge limits into waterways, and the Nuclear Regulatory Commission's radiation protection standards. The United States tends to prefer legally enforceable standards.

Guidelines, in regulation and risk management, are voluntary limits or procedures that are discretionary on the part of those agreeing to them. They are advisory, and although they may carry the weight of expectation and liability if not followed, they do not carry the weight of legal authority, and no formal sanctions are imposed if they are not followed. A guideline can become a standard if, for example, it is written into a contract or is adopted by reference as the benchmark for professional or ethical practice. Examples of guidelines include the threshold limit values (TLVs) promulgated by the American Conference of Governmental Industrial Hygienists (ACGIH) and the recommended exposure standards of the National Institute for Occupational Safety and Health. TLVs are used as voluntary standards for exposures in the workplace, and are supplemented by another class of standards called biological exposure indices (BEIs), which are guidelines for bio-monitoring (see Chapters 2 and 4). Canada tends to prefer guidelines that can be translated by provincial governments into standards, such as the Canadian national Ambient Air Quality Guidelines (formerly objectives), Canadian Drinking Water Guidelines, and remediation guidelines of the Canadian Council of Ministers of the Environment. In Canada, these national guidelines are operationalized through provincial standards.

## **Standards Setting**

Standards setting is an exercise in judgment as much as an exercise in science. The process of setting standards began in 1886 in Germany (with the maximum allowable concentration or MAK), long before formal risk assessment or risk management were developed. It began through a consensual process of some authoritative deliberating body based on a review of the evidence. Standards setting is often highly controversial, with considerable debate within the body setting or recommending standards and often suspicion of conflicts of interest and hidden agendas among the parties.

The problem is even more acute for guideline-setting organizations, such as ACGIH. The appearance of conflict of interest is particularly problematic for bodies that must, as a question of legal requirement or credibility, represent various stakeholders.

In occupational and environmental standards and guideline setting, a recurring problem is that it is very difficult and sometimes impossible to find experts on an unusual chemical or industrial process that do not have ties to the industry because that is where one learns about such things and gains real-world experience. As a practical matter and under legal pressure, standard- and guideline-setting bodies have become more transparent and accountable in recent years.

The basic approach to setting standards and guidelines is not complicated conceptually but is excruciatingly controversial in practice. It has been performed in much the same way for many years, but with increasing sophistication. The usual approach is to identify the lowest exposure levels at which health effects have been unequivocally documented in the toxicological literature (for animal studies) or the epidemiological or medical literature. This would be the level of the lowest positive finding known. If necessary, unpublished reports are used, but transparency and the regulatory process make it highly undesirable to rely on sources that are not published and peer-reviewed. A level is then identified below this level, either by the evidence or by calculation. This is a level at which no untoward effects have been observed in reasonably well-documented studies—in other words, a clearly negative study. One or more safety factors are then applied to this notional level to provide an added margin of safety. The number and size of the safety factors vary according to the uncertainties, severity of risk, and characteristics of the population involved. A larger margin of safety is used for environmental exposures, which affect many people who may be more susceptible than healthy workers, to ensure that the level is protective despite uncertainty or error in judgment or information. The presence or absence of specific susceptible groups in the population, such as asthmatics, is often taken into consideration.

This basic template still applies today, but standards setting has progressed not so much in removing the element of judgment but in defining more objectively the bounds of uncertainty. For the EPA, FDA, OSHA, or other agencies, a standard is based on well-defined toxicological measurements. Table 7.9 summarizes the toxicological measurements that are most important in setting standards using EPA terminology. The toxicological principles behind these terms are explained in Chapter 2.

The standard-setting process at the EPA is perhaps the most straightforward example that is germane to OEM. If human data are available from epidemiological or clinical studies, they are used whenever possible. The toxicological literature is searched for studies that yield a “lowest observed adverse effect level” (LOAEL) or, preferably, a “no observed adverse effect level” NOAEL (see Chapter 2). The NOAEL is then adjusted by dividing by one or (usually several) more uncertainty factors (UF) to yield a reference dose:

$$RfD = NOAEL / (UF_1 \times \dots \times UF_n)$$

The reference dose is then used as a benchmark exposure or dosage level at which the agency is confident no effect is likely to occur.

Certain practical issues typically emerge once the scientific issues are settled. These include whether a given exposure level can be reliably measured by available technology, whether control at that level is technically feasible, if so at what cost, whether a proposed safety margin provides adequate protection, whether a subgroup of susceptible workers may be at greater risk than the majority, and how the proposed standard compares with those for related agents or the same agent encountered in different settings.

## Occupational Exposure Standards and Guidelines

Occupational standards tend to be higher than environmental standards because the general population is presumed to include many ill and otherwise susceptible individuals and because, for air

**Table 7.9.** Terminology for Standards Setting and Risk Management

*Acceptable daily intake (ADI).* Used for food additives and pesticides and generally defined as NOEL/SF. Implies that a qualified authority has determined that the risk involved is acceptable, a judgment that is increasingly challenged. The ADI is used in Canadian and other standards as the basis for setting exposure levels. NOEL is “no observed effect level,” the lowest level at which any effect is observed and is used in calculating ADIs on the assumption that unless it is nutritional, any detectable effect is likely to contribute to an adverse effect in the intact animal.

*Lowest observed adverse effect level (LOAEL),* as used by EPA. The lowest level of exposure or dose that has been associated with health effects in animal studies. This measurement brackets the threshold on the high side and demonstrates that the threshold is below this level. If the NOAEL is not known, the LOAEL is sometimes used but is considered to be a poor substitute. Conventional predictive toxicology refers to LOEL, on the assumption that unless it is nutritional, any detectable effect is likely to reflect or contribute to an adverse effect in the intact animal.

*Modifying factor (MF).* This is a factor that is used to adjust the standard or RfD to take into account issues such as age or other characteristics that define susceptible groups. This term used to be part of a general safety factor. It has been split out and is considered to be a coefficient that modifies an uncertainty factor to take some potential interaction, susceptibility or further uncertainty into consideration.

*No observed (adverse) effect level (NOAEL).* The highest level of exposure or dose that has been associated with an absence of health effects in animal studies. It is always lower than the LOAEL. This measurement brackets the threshold on the low side and demonstrates that the threshold is above this level. The NOAEL can be considered a point of departure for developing a standard when the threshold is not known. Conventional predictive toxicology refers to NOEL, on the assumption that unless it is nutritional, any detectable effect is likely to reflect or contribute to an adverse effect in the intact animal.

*Reference dose (RfD),* a designation of the U.S. Environmental Protection Agency. A benchmark dose that is considered to be an estimate (within an order of magnitude or so) of a daily exposure that can be sustained by human populations, including susceptible individuals, without an appreciable

(Continued)

**Table 7.9.** (Continued)

risk of adverse effects for a lifetime. The RfD is derived from the NOEL by applying an uncertainty factor (UF), similar to SF (see below) but without implications of safety, and a modifying factor (MF) scaled from 1 to 10 reflecting professional uncertainty and completeness of the information base. In this case, the NOEL is assumed to be a no observable adverse effect level. Hence,  $RfD = NOEL / (UF \times (UF \times MF)) \dots$  as applicable).

*Safety factor (SF)*, often arbitrarily set at 100, but highly variable depending on the level of confidence that regulators have in the hazard assessment data and how severe the likely effect would be. Safety factors have been replaced by uncertainty factors, which are based on more scientific evidence.

*Threshold*. The level of exposure or dose that is associated with the earliest effect or the first appearance of an outcome. In practice, the threshold is almost never known. However, it is bracketed by the LOAEL and NOAEL.

*Uncertainty factor (UF)*. This is a factor that is used to adjust the standard or RfD to take into account statistical and biological uncertainties. This term used to be part of a general safety factor. Common uncertainty or safety factors include the following:

Uncertainty or safety factor	Adjusting for:	Magnitude
UF <sub>1</sub>	Intraspecies variation, accounting for individual differences	10
UF <sub>2</sub>	Potential interactions among agents	10
UF <sub>3</sub>	Routes of exposure when human intake is different from that used in animal experiments	10
UF <sub>4</sub>	Bioavailability or source apportionment: fraction of human intake attributable to the source being regulated	Variable
UF <sub>5</sub>	Uncertainty introduced if LOAEL is used instead of NOAEL	10
UF <sub>6</sub>	Interspecies extrapolation, applied if animal data is used instead of human	10

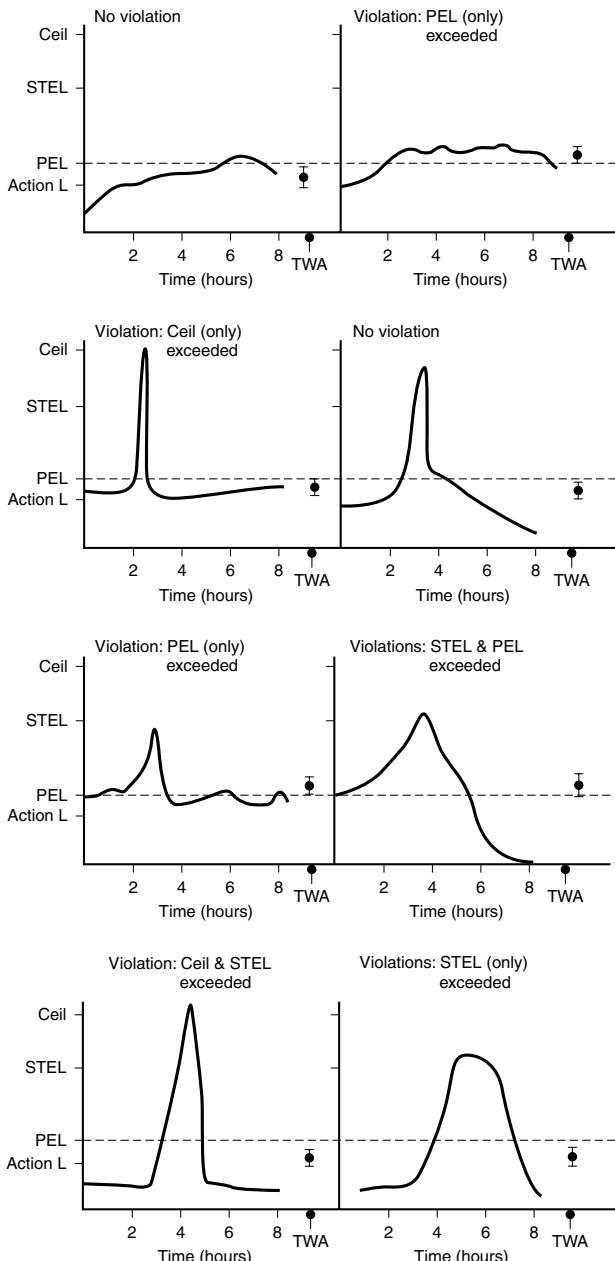
*Virtually safe dose (VSD)*. In practice, the same as the ADI but the terminology avoids the implication of an acceptable risk.

pollution and drinking water, exposure may be inescapable. In general, occupational exposure standards assume a healthy workforce containing few susceptible individuals, working a norm of eight-hour shifts. However, this assumption is increasingly out of date because of changing demographics of the working population, improved medical care, and more effective accommodation in the workplace, which has allowed greater diversity and participation by people with health conditions. The workforce is looking more and more like the general population. Many large corporations have also adopted internal corporate exposure levels for their own use, and specific standards are often proposed or recommended by industry, and professional or labor bodies, but these do not have the same force of law or authoritative recommendation behind them.

Figure 7.7 shows how standards and guidelines may be violated in different patterns of noncompliance. Risk management strategies to come back into compliance will vary depending on the reason for the exceedance (a term of art more often used in environmental monitoring) and its pattern.

## **Occupational Health and Safety Standards**

Compliance with standards may be mandatory or voluntary, the latter being equivalent to guidelines. Mandatory compliance is required for standards that carry the force of law by direct legislation or, as in most standards used in the United States, a legally prescribed regulatory procedure. In the United States, of course, the regulatory body with responsibility for setting occupational health standards is the Occupational Safety and Health Administration (OSHA), under section 6 of the Occupational Safety and Health Act of 1970 (see Chapter 25). The process of changing or setting new standards has been a lengthy and cumbersome procedure within OSHA, requiring careful and explicit documentation, formal announcement of a proposed standard, opportunities for comment from all interested parties, and often defense in court as parties who take exception have challenged the proposed standard with lawsuits.



**Figure 7.7.** Relationship of various types of standards, and possible patterns of noncompliance.

OSHA standards are known as permissible exposure limits (PELs) and are statements of the maximum concentration or level of exposure above which workers may not be exposed. PELs are usually given as time-weighted averages (TWAs)—the mean exposure averaged over an eight-hour period. In addition to PELs, various OSHA standards specify other levels. Action levels are levels below the PEL at which exposure monitoring or medical surveillance must be implemented in order to ensure that the PEL is not exceeded and that injury does not occur inadvertently; for example although the PEL for noise is 90 dB TWA, the action level is 85 dB. Short-term exposure limits (STELs) are maximum concentrations or exposure levels that cannot be sustained more than briefly, usually 15 minutes. A ceiling limit (Ceil) cannot be exceeded ever, even instantaneously. These standards are illustrated in Figure 7.5 Each standard can be exceeded individually or in combination. Ceil standards are more likely to be exceeded during accidents or failures of control equipment. When averaged over eight hours, even relatively high exposure levels, in excess of the STEL and approaching the Ceil, may be balanced by lower exposures during the rest of the work day so that the PEL may not be exceeded.

Obviously, this system has a number of intrinsic problems. It assumes an eight-hour workday, although many workers exceed this time on a work shift. It assumes, quite wrongly in many cases, that the product of duration and level of exposure define toxicity for all regulated chemicals. (See the discussion on Haber's "law" in Chapter 2.) The intrinsic toxicology of the agent may be such that short-term exposures are more harmful than longer term exposures (as has been suggested in the case of ethylene oxide) or vice versa (as has been suggested in the case of most chemical carcinogens). On the other hand, PELs may not adequately protect against agents that accumulate over time, particularly carbon monoxide, lead, and solvents. They provide limits for each agent individually but not for combinations, such as multiple exposures to similarly acting organic solvents. They are known not to be completely protective. The OSHA Noise Exposure Standard, in particular, is thought to be inadequate. The 1994 docket for review and revision of the Noise Exposure Standard con-

tains several submissions putting forward evidence that an unacceptably high frequency of workers experience noise-induced hearing loss despite compliance with the standard.

OSHA standards are difficult to reevaluate once they are in place. There are usually too many political and legal obstacles to reopen a case unless there is considerable pressure and a compelling reason to do so, as occurred in the case of benzene.

OSHA standards are now well overdue for review and revision, but the slow and encumbered pace of OSHA's approach made it impossible to respond. Instead, OSHA is likely to turn to three approaches to cut the Gordian knot: (1) the adoption of "generic" standards covering a broad range of exposures in one regulation (as was done for carcinogens not otherwise regulated); (2) the adoption *in toto*, as before, of a set of health standards proposed by, most likely, the ACGIH; and (3) adoption of standards proposed by the National Institute for Occupational Safety and Health, which has developed a well-documented set of "recommended exposure limits" (RELs). ACGIH guidelines and NIOSH recommended standards are usually more stringent than present OSHA standards. A fourth option, which would be to harmonize occupational health standards with the European Union in order to create a de facto global regulatory regime, is very unlikely for political reasons despite the obvious advantages for export to the world's largest trading bloc.

In Canada, occupational exposure limits (OELs) are set by provincial governments, which have the constitutional authority to set both environmental and occupational regulations. The role of the federal government is generally limited to overseeing the formulation of guidelines (which are developed in negotiations with the provinces) and protecting federal employees and aboriginal communities.

## **Guidelines in Occupational Health and Safety**

A voluntary or recommended standard ("voluntary standard" is a term of art in occupational health for a guideline) proposed by an authoritative body carries considerable weight. They are usually more protective than OSHA standards for the same hazards. They may be

written into a contract, influential in setting a future standard, or used as evidence of good or safe practice in a lawsuit. Employers who do not comply with generally accepted voluntary standards place themselves at risk if there is a problem, as they must be prepared to defend their departure from recommended practice if anything goes wrong or if their procedures for worker protection are challenged. Large corporations often comply voluntarily with guidelines set by NIOSH, the American National Standards Institute (ANSI), or ACGIH, and some have developed their own internal standards (really, voluntary standards or guidelines). Although enormously influential, ANSI, ACGIH, and NIOSH only propose recommended standards and have no authority to enforce compliance. The initial body of OSHA standards was adopted largely from standards current in 1970, which were recommended by either ACGIH or ANSI. (ANSI promulgates standards in many aspects of industry, including safety, but ANSI is not active in proposing health standards and will not be discussed further.)

NIOSH, of course, is a government agency, but it has no enforcement or enforceable standard-setting powers. It reviews and recommends RELs on an advisory basis, proposing changes and compiling extensive documentation on a standard-by-standard basis for recommendations that may or may not (usually not) be adopted at OSHA's discretion. NIOSH has reviewed and transmitted recommendations on many more occupational exposures than OSHA has yet acted on.

The ACGIH has been at the heart of the system of voluntary exposure standards. This is a nongovernmental organization (in spite of the name) consisting mostly of industrial hygienists, occupational physicians, toxicologists, and interested parties. ACGIH standards are called threshold limit values (TLVs), an unfortunate name because it implies that there may be a threshold of effect in a toxicological sense. TLVs are analogous to OSHA PELs in being limits, expressed as time-weighted averages over eight hours, assuming a forty-hour workweek. However, they are intended to be guidelines for repeated exposure that are acceptably safe for "nearly all" workers, day after day, to work without risk of adverse effects. Like OSHA STELs, and Ceil, there are also TLV-STELs for short-term exposure and TLV-C

(for “ceiling”) for maximum exposures that should not be exceeded at any time. In addition, the following hold true for ACGIH TLVs:

- They have the same general disadvantages as those outlined for OSHA standards.
- They presently rest on a more secure scientific footing.
- They are much more easily changed, being reviewed every two years.
- They are generally more stringent than OSHA standards and tend to be close to NIOSH recommendations. (There have been, of course, exceptions to this generalization.)
- They usually provide a greater margin of safety than current PELs, but it would be an error to assume that TLVs define levels of exposure that are demonstrably safe for all workers in all circumstances.
- Compliance with TLVs does not exclude risks; it does, however, minimize the likelihood of an unanticipated adverse effect.

ACGIH has developed standards that have been used as guidance or even adopted without change as mandatory standards in many other countries, particularly at the provincial level in Canada, Europe prior to the formation of the European Union (EU), Israel, Japan, and many developing countries. (Historically, Soviet standards were more influential in Eastern Europe, except for the former Yugoslavia.) Until recently, relatively few countries, among them China, have attempted to develop a completely or largely independent standard-setting capability. Now the European Union is the world leader with a comprehensive body of standards for worker protection. Because of the importance of the European market and the influence of the European model for standards setting, more countries are now adopting European standards or are influenced by the EU process, which has increasingly come to supplant ACGIH on the international scene.

ACGIH has supplemented its TLVs with a set of biological exposure indices (BEIs, see Chapter 4), providing biomarkers and indica-

tors of exposure that are adapted from toxicological testing or biological-effect monitoring. They are discussed in the next section.

As a private, professional organization, ACGIH wields enormous influence, and with it comes scrutiny and the need for accountability. The scientific basis and objectivity of ACGIH standards have been called into question, particularly in Germany and by critics in the United States. In part, this reflects ACGIH's role as an innovator, having stepped into a vacuum in 1938 when local governmental authority did not support the establishment of standards and the only federal presence had to do with research. ACGIH established a committee to create recommended standards and to collate all available information, much of it directly from industry sources. At the time, there was very little documentation in the open scientific literature. Over the succeeding years, the process became rather elaborate. It came to be increasingly dominated by industry representatives or consultants who may have had a conflict of interest and who certainly had access to more extensive resources than those who might have argued for more stringent standards. In recent years, much of ACGIH's work has been questioned and there have been allegations of undue industry influence. Given the circumstances of the time, however, ACGIH was highly restricted in access to information and relied heavily on knowledgeable consultants. The professionals most knowledgeable about hazards in a particular industry invariably have ties to that industry, precisely because that is normally how one obtains this expertise. Although cupidity is not unknown in professional affairs, many of ACGIH's lapses seem to reflect the standards of the times and the constraints of their work. The organization has made strenuous efforts to be objective and transparent and in recent years has faced its strongest opposition from industry, in the form of legal action to block promulgation of TLVs. Whatever the past, ACGIH's TLVs and BEIs will continue to require periodic and comprehensive review in light of new scientific information and insight.

There are many fundamental problems, as well as strengths, in the process of setting mandatory or voluntary standards as it is conducted in the United States. One is that the selection of a standard is a decision

based on social consensus regarding what represents an acceptable risk, since complete safety cannot be assured at any level of use of a chemical. Authoritative professional organizations are presumably best able to perform the risk assessment function but are not necessarily suited to the risk management function that follows. This is in part because these bodies lack much of the expertise needed but more because in the process of rendering the analysis more vigorous, objective, and substantiated by data, they tend to submerge or deny the uncertainty, social issues, and cultural values that are implicit in the decision. On the other hand, attempts to deal with standards setting by legislation or in any other forum of political give-and-take and negotiation carries the risk of corrupting science and compromising the essential interests of less powerful or articulate parties—in this case, exposed workers. There is always pressure to politicize the issue and for participants to play roles derived from their experience, self-identification, and those interests that support them. A strength of the OSHA process, cumbersome as it surely is, is that the political and social context of decision-making is made explicit, if not always transparent. It remains for future reform to devise a means of making it responsive and expeditious.

In the meantime, as noted, the worker protection standards of the European Union are becoming the new world standard. As well, a new strategy called control banding has been developed to guide hazard control without requiring extensive exposure monitoring. (Control banding is described in Chapter 14 with hazard control.)

Standards and guidelines are not generally useful for the prevention of allergy because they bear no relationship to sensitization thresholds. Individuals who become sensitized to a chemical may experience allergic responses at or far below even the most stringent standards of exposure.

### **Acute Exposure Guidelines**

Since 1986, the EPA has developed a set of guidelines for acute exposure of the public in emergency situations called “Acute Exposure Guideline Levels” (AEGLs, pronounced “eagles”). These standards are

intended to guide emergency management during extraordinary events in which exposure is unavoidable during rescue or response. The operating assumption is that a substantial fraction of the general population would be susceptible, such as infants, those suffering from chronic disorders such as asthma, or the elderly, but that such exposures would only occur once in lifetime. AEGLs are set by a national advisory board that has representation from the private sector.

AEGLs are set for five time periods (10 minutes, 30 minutes, 60 minutes, 4 hours, and 8 hours) and three levels. The three levels are as follows, with their standard descriptions:

1. AEGL-1 is the airborne concentration (in either parts per million [ppm] or mg/m<sup>3</sup>) of a substance above which it is predicted that the general population, including susceptible individuals, could experience notable discomfort, irritation, or certain asymptomatic nonsensory effects. However, the effects are not disabling and are transient and reversible upon cessation of exposure.
2. AEGL-2 is the airborne concentration (ppm or mg/m<sup>3</sup>) of a substance above which it is predicted that the general population, including susceptible individuals, could experience irreversible or other serious, long-lasting adverse health effects or an impaired ability to escape.
3. AEGL-3 is the airborne concentration (ppm or mg/m<sup>3</sup>) of a substance above which it is predicted that the general population, including susceptible individuals, could experience life-threatening health effects or death.

Obviously the objective is to make every possible effort to keep exposure below any of these levels, but the guidelines provide a benchmark for what to expect. Table 7.10 demonstrates the use of AEGLs for hydrogen sulfide, a highly toxic gas in which the initial concentration is much more significant in determining toxicity than the duration of exposure.

**Table 7.10.** EPA Acute Exposure Guideline Levels for Hydrogen Sulfide (in ppm), for which the Odor Threshold is 0.01 ppm.

AEGL Level	10 minutes	30 minutes	60 minutes	4 hours	8 hours
AEGL 1	0.75	0.6	0.51	0.36	0.33
AEGL 2	41	32	27	20	17
AEGL 3	76	59	50	37	31

## BIOMONITORING PROGRAMS

Biological exposure indices (BEIs) are measurements of concentrations of chemicals (xenobiotics) or their metabolites in body fluids or expired air or indirect measurements based on the effects of the exposure. They are biomarkers of exposure (see Chapter 2) that have been developed by ACGIH as guidelines for individual exposure monitoring. They are the counterpart for biological monitoring to the TLV, which is a voluntary standard for environmental exposure in the workplace. BEIs have been developed for a variety of potential exposures, based on a knowledge of toxicokinetics, practical measurement technology, and assessment of intra- and inter-individual variation for purposes of standardization.

The measurement methods of a BEI reflect the internal dose, not environmental exposure. Workers may be exposed to the same chemical (such as solvents) at work and at home or in avocational pursuits such as hobbies.

The strategy for taking measurements (usually postshift) ensures that occupational exposure dominates in the event of multiple sources. The timing depends on the toxicokinetics of the agent of concern (see Chapter 2). Volatile organic compounds should be measured in expired air immediately post shift. Blood lead levels represent cumulative, not recent, exposure and should be monitored periodically and over a period of weeks, not days. Serum cholinesterase levels for organophosphate exposure respond quickly, but the exposure is intermittent, not tied to a shift, and so should be measured at baseline before any exposure occurs and then after significant exposure and in the event of

symptoms. For reasons of compliance, occupational justice, and legal requirements, sample collection for BEIs is done on the employer's time, not on the worker's time after the workday has ended.

BEIs cannot be used when the xenobiotic or its metabolite are also a product of normal metabolism at relatively high concentrations. The body produces carbon monoxide endogenously from bile metabolism, so there is some baseline value of carboxyhemoglobin for all people. If it were higher, carboxyhemoglobin or expired air carbon monoxide could not be used. BEIs are not helpful when the adverse effects are local or superficial, such as skin or mucosal irritation or bronchial inflammation. They are not useful with respect to allergy since sensitization thresholds are not related to what they measure.

BEIs set levels of personal exposure that should be nearly universally protective; in other words, the subpopulation of workers who remain at risk at these low levels should be very small. BEIs have been relatively slow to replace TLVs but have many advantages in monitoring individual exposure. BEIs are exposure levels for the individual that reflect absorbed dose from all sources and reflect individual characteristics of the worker (such as size, adiposity, and metabolic capacity). They are based on the toxicokinetics of the hazard and assume that the worker is both healthy and has a normal capacity for biotransformation of those chemicals that are metabolized. They are used as guidance for the individual worker.

BEIs and other biomonitoring methods have great potential for advances in the management of individual as well as group risk. However, they are not without problems in implementation.

Biomonitoring programs have not been uniformly welcomed into the workplace. They are often opposed by workers on grounds of an invasion of privacy and person. They can be easily confused with drug screening programs, and workers usually suspect that there will also be testing for alcohol and nicotine products (cotinine). Workers are legitimately suspicious of anything that suggests genetic testing and are concerned that disclosure of results may affect their future insurability, especially if it uncovers a pre-existing condition or disease risk. They may be concerned that a label of ill health will be applied to

them and be afraid of stigmatization, particularly if they do not trust safeguards for confidentiality by the employer.

Programs for biomonitoring should be handled carefully and with an emphasis on communication with workers and slow introduction into the workplace. The legal, ethical, and labor-management issues should be worked out well in advance. The rights of those being monitored should be respected. It is fair to make participation in narrowly focused biomonitoring programs for BEIs a condition of employment, but it can be difficult to make this retroactive for current employees. On the other hand, new hires may feel that a policy that exempts veteran workers is discriminatory. It is not acceptable to introduce programs for genetic screening or for any characteristic that links to race or ethnicity.

Before initiating a biomonitoring program, the OEM physician should clarify why it is being done and why it is preferable to conventional environmental monitoring. There should be clear goals for what is to be accomplished. The logistics for implementation should be worked out and pretested. Only established, well-accepted methods should ever be used for the analysis procedure. (Innovative or new methods risk error through laboratory unfamiliarity and anomalous, uninterpretable, or questionable results.) Documentation should be thorough, and the worker's records should be available to him or her for inspection.

The goals of a biomonitoring program should be to protect the worker. It should be used to identify individuals at unusual risk because of exposure to a hazard before they become ill and to deal with the problem appropriately without penalizing the worker. It should be used to identify workplaces and jobs where control of hazards is not effective so that this can be corrected. Under no circumstances should monitoring programs, biomonitoring or environmental, be used as a substitute for controlling the hazard in the first place.

## **RESOURCES**

Lundgren R, McMakin A. *Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks*. Hoboken NJ: Wiley IEEE Press; 2008.

- Robson MG, Toscano WA, eds. *Risk Assessment for Environmental Health*. San Francisco: Jossey Bass/John Wiley; 2007.
- Slovic P. Perception of risk. *Science*. 1987; 236:280-285. [Classic, must be read.]

### **NOTEWORTHY READINGS**

- Calabrese EJ, Gilbert CE. Lack of total independence of uncertainty factors (UFs): implications for the size of the total uncertainty factor. *Regul Toxicol Pharmacol*. 1993;17(44):44-51.
- Douglas M, Wildavsky A. *Risk and Culture*. Berkeley: University of California Press; 1982. [Classic]
- Graham JD, Wiener JB. *Risk vs. Risk: Tradeoffs in Protecting Health and the Environment*. Cambridge: Harvard University Press; 1995.
- Leiss W, Chociolko C. *Risk and Responsibility*. Montreal: McGill-Queens University Press; 1994.
- Lowrance WW. *Of Acceptable Risk: Science and the Determination of Safety*. Los Altos: William Kaufman; 1976. [Classic]
- McGraw TK. *Prophets of Regulation*. Cambridge: Harvard University Press; 1988. [Classic study of key figures in the development of government regulatory policy.]
- National Research Council. *Risk Assessment in the Federal Government: Understanding the Process*. Washington, DC: National Academy Press; 1983. [The original Red Book. A revision is in preparation.]

*This page intentionally left blank*

# **8 PREVENTION SCIENCE**

Prevention science, comprising the tools of preventive medicine, is simply a systematic approach to wellness. Occupational and environmental medicine (OEM) has a fundamental mission to prevent illness and injury arising out of exposures at work and in the environment and a secondary mission to promote a healthy workforce and healthy community. OEM is considered a prevention-oriented medical specialty, and its specialty board (in occupational medicine, specifically) in the United States is the American Board of Preventive Medicine, which also certifies specialists in general preventive medicine and aerospace medicine. Wellness is already an important part of occupational medical practice. The packaging of prevention services into wellness programs is discussed in detail in Chapter 19.

The scientific basis for preventive services is laid out in detail in textbooks of clinical epidemiology or public health. Much of the science of prevention is based on epidemiology (which is discussed as a foundational science of occupational and environmental medicine in Chapter 3). This chapter is intended to lay out only the basic concepts and to make the connection with OEM.

The role of physicians today in all primary and secondary care specialties is to maintain good health as well as take care of the sick. One

might think that preventive medicine is the opposite of clinical medicine and the treatment of disease. However, much of clinical practice is already devoted to preventing disease, to ensuring that complications do not occur, and to preventing disability from arising out of illness and injury. Prevention is an integral part of medical practice, part of the physician's mission to sustain health, relieve pain and suffering, and treat disease, preferably before it starts.

As the health of patients and the community continues to improve in youth but to be threatened by age as people live longer, there is a greater emphasis on prevention of disability. Physicians find themselves increasingly charged with providing care to keep people healthy rather than to treat them for illness. If one can do a better job in preserving health, then presumably the system as a whole will do a better job in taking care of the ill, because it will be dealing with people who are ill less frequently and less severely ill when they do become ill. In financial and managed care terms, therefore, prevention equates to demand reduction for the system as a whole and is to be encouraged. Yet another benefit is the importance of preventing disability and premature illness, and the resulting withdrawal from the workforce, in maintaining economic productivity in an aging population.

Despite the promise of prevention to achieve health gains, the personal benefits of wellness to the individual, and the benefits to society of sustained productivity, prevention lags far behind treatment-oriented medicine in terms of funding commitment, professional and public interest, insurance reimbursement, public policy, and medical education. The seemingly less urgent healthcare services of prevention are easy to postpone, delay, or forget, especially under the pressure of a busy practice, a crowded medical curriculum, and a busy life. Much is left undone. Compliance on the part of patients and clients leaves much to be desired. In part, this is because most physicians and health policy advocates do not generally understand prevention beyond a superficial level. It may also be that the popularization of prevention science has attracted advocates of unproven medical theories and that the risks uncovered by epidemiology have been so confusingly presented and contradicted

in media reports that healthcare professionals have come to distrust the field. Among biomedical scientists, prevention science and epidemiology are still widely perceived, and sometimes derided, as “soft sciences”—as if humanitarian concern were a mark of weakness, social and behavioral change were straightforward, and the natural constraints of a scientific discipline reflected on the moral weight of the investigator.

However, commitment to prevention is lacking mainly because the current healthcare system does not actively support it. Education, counseling, vaccination, and most clinical screening services are poorly reimbursed and carry few other incentives. Although some managed care organizations, such as Kaiser, have made progress, the healthcare system overall is only now beginning to find effective formats for incorporating prevention into clinical practice routinely in an effective and sustainable way without financial penalty. There is also a strong opinion among economists that effective prevention actually increases the cost of healthcare over the long term by reducing premature mortality and, as a consequence, improving survival in people with disability and chronic, expensive health conditions. Prevention advocates respond—passionately—that health is a personal and social good in itself, well worthy of investment, and that effective prevention also prevents disability and keeps people productive longer. In recent years there has been increased emphasis on the prevention of disability (“tertiary prevention,” explained below) with the objectives of obtaining better outcomes from treatment and rehabilitation, reducing the cost of chronic care, and achieving the highest level of function as a person ages. From an economic point of view, this is important for the aging populations of developed countries and to prevent a long-term social burden from impeding the progress of developing countries. From a social point of view, it is critical to empowering people to live the best lives they can and to contribute what they are able, regardless of disability. (See Chapter 18.)

The technology of preventive medicine is not complicated. The rationale and the information base underlying preventive medicine,

however, are based on a theoretical framework that can become quite elaborate.

## **THE BASICS OF PREVENTION**

The fundamentals of preventive medicine are not complicated. Implementation, however, can be very complicated and challenging, not least because many preventive measures depend on changing individual behavior.

Preventive interventions are classified on three levels: primary, secondary, or tertiary, depending, respectively, on whether they prevent the occurrence of disease by reducing exposure or modifying risk factors, detect disease or a marker of risk early enough for successful intervention, or limit available disability once a disease has occurred. These strategies are complementary, not competitive. Application of these levels of prevention rests on two traditional modes of disease prevention. An attempt to change the determinants of risk to the entire population is the public health strategy. An attempt to identify and control determinants of increased risk for a subset of individuals who are at unusually high risk is the (clinical) preventive medicine strategy.

Prevention is incorporated into broader programs in which the emphasis is on primary prevention to reduce disease incidence in the working population and on secondary prevention, to detect disease early and to refer patients for care when intervention is most likely to be effective. The preventive medicine approach in the workplace, directly applied, is typically limited to screening for common disorders and risk factors and intervention activities that supplement but do not substitute for personal healthcare. Reducing risk factors for later health problems (such as reducing cholesterol levels through cardiovascular fitness training) is more easily accomplished in an integrated health promotion program. In such programs peer group influence, constant encouragement and feedback, and support networks make compliance easier to achieve compared with solitary individual interventions. Most worksite wellness programs appropriately emphasize multiple rather than single risk factor reduction.

## What Is Health?

Health is a surprisingly elusive concept, defined for the most part by medical and other healthcare practitioners as the absence of what it is not: illness. Efforts to define health as a form of well-being have tended to struggle with operational boundaries among physical health, mental health, personal well-being, and community integration. The most widely accepted definition of health is that of the World Health Organization: “not just the absence of disease but a complete state of physical, mental and social well-being.” This definition has been hard to operationalize because it is so general.

This, and any, definition of health must operate on two levels: personal and communal. In the communal perspective, each individual may be sick, well, or functional at various levels, but it is the collective, the average, or the distribution of levels of health status that can be monitored and that is therefore of concern to public health agencies and government. This communal level of health is often called “population health status.”

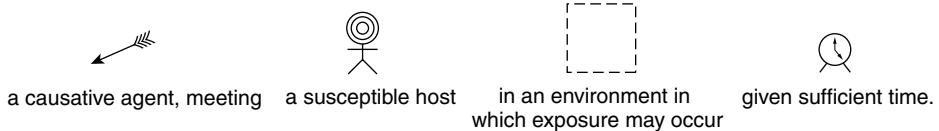
Further discussion of the philosophy of what constitutes health, while fascinating, is beyond the scope of this book. Use of the WHO definition will be assumed throughout.

## Levels of Prevention

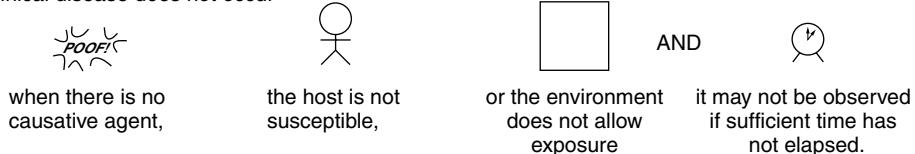
There are three traditional levels in preventive medicine (Figure 8.1):

- Primary prevention is the first, and most fundamental, level. It is the prevention of the occurrence of the disorder, usually by preventing exposure to the agent that causes it or by preventing the expression of the disease by enhancing the resistance of the host, as in the case of immunization to prevent infectious illness.
- Secondary prevention is the prevention of the progression of a disorder by early detection and intervention. If possible, one should identify a disorder in the earliest stages possible so that it can be treated with the highest probability of cure or arrest.

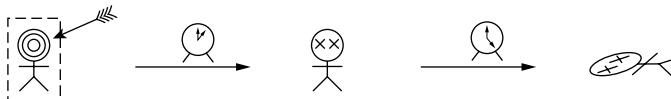
Clinical disease results from



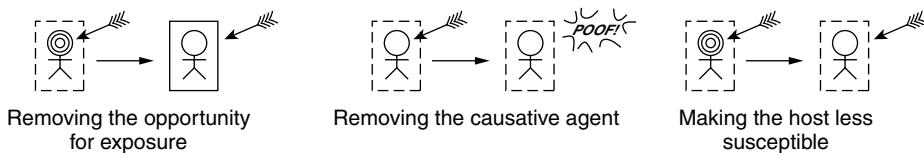
Clinical disease does not occur



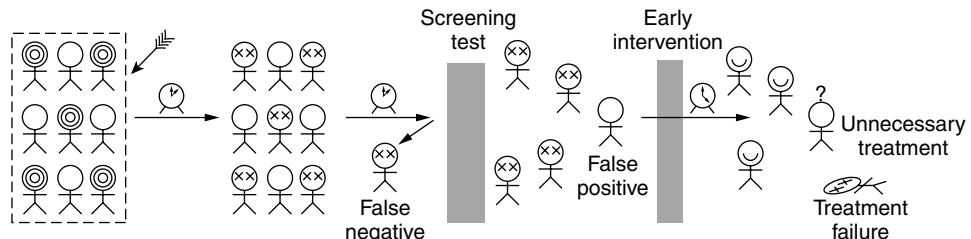
Many diseases also have subclinical, early phases lasting for periods long enough to make detection feasible.



Primary prevention works by doing one or more of the following:



Secondary prevention works by screening to identify subjects with early disease and treating them:



**Figure 8.1.** Levels of prevention.

- Tertiary prevention is the prevention of disability or complications resulting from a disorder through rehabilitation and specific treatment intended to prevent complications. Tertiary prevention is often neglected in preventive medicine but has the potential to make a greater contribution as the population ages.

The same approach of primary and secondary prevention of disease can be applied to preventing risk factors for disease. Health behaviors (such as unsafe sexual behavior), risk factors (such as elevated cholesterol), environmental exposures (such as second-hand smoke in the home), and social determinants of health (such as family violence) can be managed as if they were diseases using the same framework. For example, in smoking cessation we can think of smoking as an outcome and we can consider the “causes” and risk factors of smoking. Some conditions are both diseases and risk factors, such as hypertension, and represent both outcomes in themselves and risk factors for other disorders. Identification of persons at risk on a screening test or interview allows the physician to intervene to change that risk factor and prevent the outcome. Thus, the methodology of secondary prevention can also be used to support primary prevention.

### ***Primary Prevention***

The logic underlying prevention begins from simple building blocks. In the case of primary prevention in OEM, the approach is largely a matter of identifying the cause and preventing exposure. Another alternative is to modify the host to reduce susceptibility to disease. In the case of immunization, the approach involves modifying the host’s response to the disease agent by giving the host a vaccine that confers immunity. There are fewer opportunities to do this in the workplace, but it is a cornerstone of general prevention.

A more complex situation arises in primary prevention based on changing health-related behavior. The goal is to change the risk factor or the health-related behavior in order to prevent the target

condition rather than directly manipulating the potential for exposure to the cause. For example, in the case of cigarette smoking, it is possible to change people's behavior so that they will reduce consumption despite the continued availability of tobacco products. In OEM, the goal is usually to avoid relying on behavioral change in workplace settings, but behavior change is critical for reducing exposure to lifestyle risk factors. This relatively simple step to changing behavior has profound consequences because it immediately thrusts medicine into the social and psychological realm and outside of pure biology.

In primary prevention, it is often difficult to link a change in a risk factor with the eventual clinical outcome, especially if there is a long delay, as is typically the case in cancer prevention. Instead, investigators break the process down into pieces and look at each piece in turn. First, one identifies the probable cause and considers how exposure to the risk factor can be modified or avoided through behavior. One must look as well at intermediate points to monitor progress in changing exposure to the risk factor. In the case of osteoporosis, for example, one may look at whether progressive bone loss is prevented by supplemental calcium, vitamin D, estrogens, or fluoride as an intermediate step, but it will take many years to be sure that clinical osteoporosis can be prevented. It is not practical to track the success of the intervention all the way from compliance with taking the supplemental medication to the end point of preventing clinical osteoporosis within a reasonable time frame. One has to identify these intermediate points on the way to a complete evaluation: how to get from point A to point B and then from point B to point C. In this way, recommendations for preventive practices are usually evaluated piecemeal rather than at once in one clear, well-defined study.

Primary prevention requires simple interventions that are sustainable, have low risk, and are cost-effective. Otherwise, the payer (whether an insurer, employer, public health department, or health-care institution) will not continue them and the benefit will be lost. Evaluating primary prevention programs is anything but simple, however, because it necessarily involves, at the very least, knowing

the etiology of the condition, knowing the efficacy of the intervention, assessing the effectiveness of the intervention in practice, determining the risk associated with the intervention (for example, side effects of a prophylactic drug or vaccine), evaluating the performance of programs in delivering the intervention, assessing levels of compliance with the intervention, determining obstacles to adopting or complying with the intervention, and assessing the cost-effectiveness, risk-benefit, and return on investment (compared to other health interventions).

The ethical issues in primary prevention are not simple either. The fundamental ethical principle of medicine is “First, do no harm” (*Pri-mum non nocere*). Because an intervention is being given to people without a problem (not necessarily healthy in other respects but at least without the disorder being prevented), it is unacceptable for the intervention to carry a substantial risk. In deciding whether to accept a treatment, the patient (with the guidance of the physician) balances the risk of the treatment against both the risk of progression of the disease and the potential benefit to be gained from the treatment and may accept a high risk of side effects if the disorder is life-threatening, painful, or dreaded. In contrast, an intervention in primary prevention must always have a very low risk-benefit ratio. That is why vaccines with a risk of serious side effects of 1 in 100,000 are unacceptable for the general public and why concern over low levels of disinfection by-products (trihalomethane compounds, which are potentially carcinogenic) in drinking water has led to measures to control chlorine levels in disinfection, despite the extremely low risk. Because in many circumstances vaccination is compulsory (admission to school, military) and hazard control (for example, treatment of drinking water) is essentially compulsory because it is difficult to opt out (unless a person has his or her own water supply), there is an additional ethical dimension of coercion, notwithstanding the collective benefit. The gains to every member of the population from immunization against a communicable disease are much greater because of the increased protection that derives substantially from group, or “herd,” immunity, so the exercise of one’s right to refuse

immunization is potentially threatening to everyone else. The balance between the interests of the individual and that of the collective, never settled in North American society, has often been challenged by activists who isolate this element of benign coercion for the public good and turn it into an ideological test. This propensity to separate primary preventive measures from their context motivates and underlies many health issues, as in the case of allegations of a relationship between childhood immunization and autism (which the evidence does not support), the anti-fluoridation movement, and groups opposed to pasteurization of cheese and milk (which in raw form can transmit serious infections). Clinical prevention services for adults, on the other hand, are mostly voluntary and usually do not raise these issues. An adult can decline a flu shot and take his or her chances of getting the flu. A child will not be able to go to school without being up to date on childhood immunizations. When many children fail to be vaccinated against a disease such as measles, they not only risk becoming victims of an outbreak themselves but also expose others.

Most occupational, environmental, and public health measures are on the level of primary prevention: clean water, safe food, pollution control, and other measures designed to prevent exposure in the first place. Primary prevention approaches must be based on sound etiologic research, and OEM physicians in academic and government research institutes (such as NIOSH) are heavily engaged in such research. In OEM, primary prevention is mostly about hazard control (see Chapter 14), which falls more in the domain of occupational hygienists and process engineers on the occupational side and to environmental engineers, environmental specialists, and environmental health professionals (including health inspectors) on the environmental side. The role of the OEM physician in primary prevention may be greater than in other medical specialties and is central to research in OEM, but there is a practical limit to what a physician in practice can do in managing the workplace or ambient environment. On the other hand, primary prevention also involves rendering the host less susceptible to disease through measures such

as immunization. OEM physicians do this in many ways, both direct and indirect, such as performing respirator fit testing to ensure that personal protection can be used.

### ***Secondary Prevention***

Historically, much of OEM involved secondary prevention, particularly occupational medicine. Even today, periodic health surveillance for the early detection of occupational disease remains an important part of OEM practice, driven in part by OSHA standards that require periodic health surveillance. However, secondary prevention as the main strategy for worker protection is a diminishing part of OEM practice in North America. In France, where the occupational medicine system is more heavily oriented to prevention, it remains the primary function of occupational physicians. Environmental medicine has always relied heavily on secondary prevention—for example, blood lead screening for children.

In secondary prevention, the process is complicated by the need for a screening step. This screening step is necessary in order to identify individuals at risk or individuals who have the disease in an early stage. To do so requires a clinical service or test. An early detection procedure identifies a target condition or risk that can be modified by a treatment or other intervention to improve the chances of a favorable outcome. A secondary screening strategy involves an additional complication because the screening tests selected must have certain characteristics. Ideally, the test should be very sensitive, specific, simple, cheap, and safe. Safety is a particular concern in order to avoid causing more harm than good by subjecting a large number of normal people to a small but real risk. The early detection procedure itself must be evaluated for validity and operational feasibility, and that can be a very complex task. This is just the beginning of the evaluation.

Like primary prevention, secondary prevention must be grounded on a firm knowledge of etiology and the natural history of the disorder being detected, especially in its early, even asymptomatic stages.

Unlike primary prevention, secondary prevention must also contend with the sensitivity, specificity, and predictive value of available screening tests, which add another level of complexity and uncertainty. The test must also be acceptable to the subjects and to those conducting the test. It cannot be too inconvenient, too time-consuming, too uncomfortable, or too unpleasant. Program planning and evaluation must take into account the performance of the test in the specific population to which it is applied (workers or the community, which may have very different prevalence and therefore predictive value for the test, as described in Chapter 5), the performance and risks associated with confirmatory tests (because screening tests are usually not definitive), and the cost and feasibility of delivery of the test. (It is not practical to offer colonoscopy at a health fair, for example.) By these standards some tests, such as stool guaiac for colon cancer, are worthwhile even though they are highly insensitive, but many sophisticated tests, such as periodic surveillance for lung cancer and PSA (prostate-specific antigen) for early detection of prostate cancer, are not.

Once the screening tests are evaluated, the efficacy and safety of various interventions have to be considered. This leads to a much more complex schema than in the case of primary prevention and opens the possibility of different ways of getting to the same point. One might choose to study several detection methods, for example, as well as several interventions. When the time comes to fully evaluate all of the possibilities, the problem becomes very complicated indeed. Our data are hardly ever complete enough to really know what the best course is. Unfortunately, the information on which we base important prevention-oriented actions is almost never complete or even fully satisfactory. As the epidemiologist pioneer Dr. Alexander Langmuir once said, "The data you want you don't have, the data you have you don't need, and the data you need probably don't exist."

Obviously, screening is not appropriate as an individual strategy for health protection if there is no treatment that modifies the outcome. However, screening may identify a pattern of disease risk, which may

lead to protection of other workers. This is the use of periodic health surveillance to guide primary prevention, not secondary prevention.

OSHA standards require periodic health surveillance for workers exposed to certain hazards (such as asbestos and noise), and employers make compliance with the standards a requirement for the job, so this form of secondary prevention is largely compulsory. However, in the community and for personal health promotion, secondary prevention is voluntary. Effective prevention must be organized and marketed in an attractive and convenient program that motivates people to comply and also to follow up on the screening test results.

### ***Tertiary Prevention***

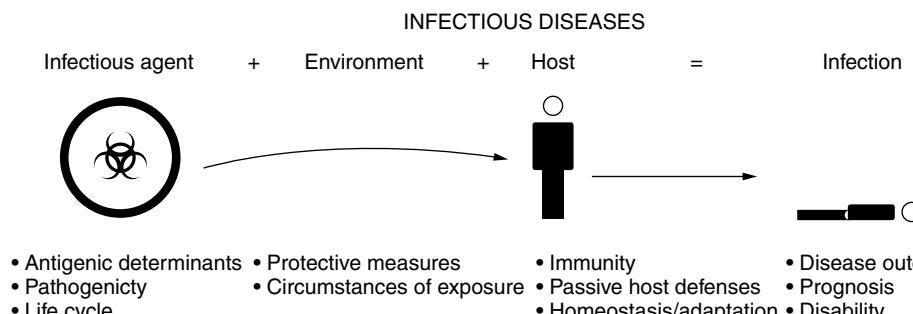
Tertiary prevention, which is the prevention of disease progression and disability, is particularly germane to the workplace because it protects the workers' employment and social and future earnings capacity. Tertiary prevention also fits well with the goals of workers' compensation and disability management. This topic is discussed further in Chapter 18, on work capacity. The prevention of disability and of complications of chronic disease is playing an increasing role in the health-care agenda in North America and is compatible with the drive toward "better outcomes" and "faster cures" in clinical medicine.

### **The Public Health Triad**

The public health triad is a concept in primary prevention as it applies to infectious disease and environmental and occupational health. It was first developed as a conceptual framework for tropical medicine.

For an infectious disease to occur, there must be three elements present (Figure 8.2):

- The pathogen
- An environment that allows exposure to occur
- A susceptible host



**Figure 8.2.** The public health triad for infectious disease: infectious agent, environment, and host.

The pathogen is an infectious agent with sufficient pathogenicity to cause the disease under ideal conditions. The environment must bring the pathogen and the susceptible host together in some way, for example by supporting the growth of vectors such as mosquitoes, through water contamination, or some other pathway that brings the pathogen and the host in contact. The host has to be susceptible to the pathogen. This means that the host is not immune, that passive host defenses are not adequate to prevent infection, and that symptomatic disease can occur in the host. The three together are required for the disease to occur.

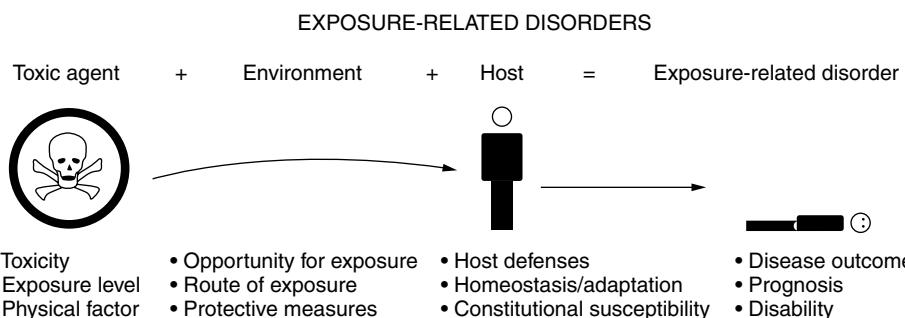
Interrupt any one of the elements and the host is protected against the disease. If the pathogen can be eliminated, as in the case of cholera, disease cannot occur and the entire community is protected, but there is always a risk of reintroduction. If the environment is changed, for example to control vectors or to supply clean water, the pathway of exposure is interrupted. The pathogen persists, however, so interrupting the pathway of exposure often depends on continued monitoring and vigilance, but it can be done for everyone in the community at once. The host can also be modified, most obviously by immunization or by drug prophylaxis (for example, by antimalarial medication). Managing the host requires that each individual receive the intervention and everyone entering the community who is not protected or who is born into the population must receive the intervention to be

protected. Therefore a commitment to modifying the host is a long-term commitment to providing preventive medical services.

Although it is not obvious, chemical exposures and other hazards function in a similar manner. For disease related to a toxic or physical hazard to occur, there must likewise be three elements present (Figure 8.3):

- The toxic agent
- An environment that allows exposure to occur
- A susceptible host

A toxic agent, for example, has to be intrinsically toxic to be a threat, and the exposure level has to be sufficient to cause toxicity. The environment must provide a pathway of exposure that brings the hazard and the host into contact, such as air pollution, water contamination, or skin contact in the workplace. For environmental hazards, demonstrating an “intact pathway of exposure” is critical to the investigation of the risk of toxic chemicals. The host has to be susceptible for disease to result; as discussed in Chapter 2, there is variability in susceptibility to chemical hazards, in particular, and some conditions, such as atopy and asthma, are associated with a much higher risk of certain outcomes, such as mucosal irritation and airways reactivity.



**Figure 8.3.** The public health triad for chemical and physical hazards, illustrated with the former: toxic agent, environment (and intact pathway of exposure), and host.

Interrupt any one of the elements and as with infectious disease, the host is protected against the disease caused by these hazards. This can be achieved by eliminating the hazard or controlling exposure, which in environmental health is often done by regulation. The environment can be changed so that exposure does not take place or is reduced, and there are many ways to do this in OEM (see Chapter 14). The host can be modified, broadly, by personal protective equipment that greatly reduces risk by preventing exposure on a personal level, although this does not necessarily protect others.

Haddon's Matrix (discussed in Chapter 6) is an application of the public health triad idea to injuries. Haddon had the insight to apply it to kinetic energy in the abstract. The other category of occupational hazard, psychosocial stressors, is more complicated but also has elements of the public health triad implicit in the idea of stressors, the social dimension of stress, and resilience.

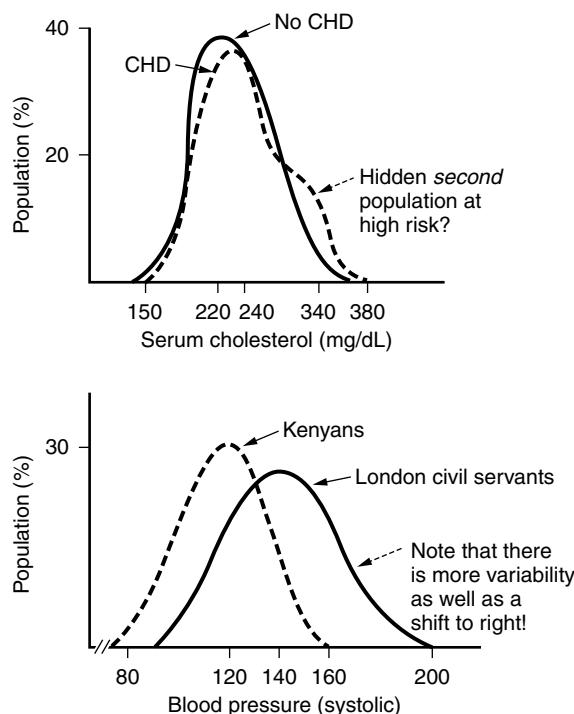
The value of the public health triad is that it is a framework for thinking about infectious and environmental health problems, including occupational health issues. It offers three robust conceptual approaches to solving every problem related to exposure and risk.

### **Strategies for Prevention**

There are two basic strategies that are used to change the individual risks of having a disorder. These differences define the approaches of public health compared to preventive medicine. One is to look at those individuals who are at high risk because of exposure to a cause of the disease caused by their personal characteristics and concentrate on them. The other approach to take is to look at the whole population and its distribution of risk factors as a whole and treat all of them, whether they are at elevated risk or not. The “high-risk” or preventive medicine strategy requires a screening step, either to identify persons who are developing the condition (the secondary prevention approach) or to identify persons with behaviors or risk factors that can be modified by an intervention to protect them from getting the disorder in the first place (the primary prevention approach).

The distinction between the “preventive medicine” and the “public health” approach was first suggested by Sir Geoffrey Rose, a distinguished British epidemiologist who in the 1980s developed his theories based on the example of cardiovascular disease. Figure 8.4 illustrates two of Rose’s most famous examples.

Rose addressed the question of whether efforts taken to reduce the risk of persons who were not at high risk were worth the effort (see Figure 8.4, upper panel). He pointed out that in the case of cardiovascular disease, in particular, the great majority of cases that come not from the high-risk groups, but from people in the normal risk groups. Who will eventually develop disease is not obvious because serum cholesterol is not the sole determinant of heart disease, and



**Figure 8.4.** Examples (redrawn) used by Geoffrey Rose to illustrate the “public health approach” and the “preventive medicine approach” to prevention.

therefore some individuals with cholesterol levels below the target screening level will go on to develop disease anyway. Therefore to intervene on a population level to reduce the risk of the entire population makes a good deal of epidemiological sense in this case because most of the cases do come from lower risk groups, even though the people who have the highest level risk factors are the ones who have the highest risk for the disease. This is called the “Rose paradox,” and it holds true for many diseases. This approach to intervention Rose called the public health model.

Rose also used the example of hypertension, in which one might look at distributions of systolic blood pressure in two populations: residents of Kenya (where mean blood pressure is low and hypertension rare) and civil servants in London (where mean blood pressure is higher and hypertension is common; Figure 8.4, lower panel). In the case of a population of London civil servants, the curve is shifted compared to individuals in an African society, in which the mean blood pressure is substantially lower. The mean is displaced although the distribution remains similar but somewhat wider. One could try to change the entire London population so that it matches that of an extremely low-risk African population by, say, dietary intervention restricting salt or even pharmacologic intervention. Many people would be treated who would not require treatment, and some might develop side effects from treatment. However, the complications of hypertension would be prevented in all people, and many cases in the low-risk groups would be prevented. The alternative is to consider people who are above a certain arbitrary level as being at high risk. A blood pressure screening program would identify them and direct them toward a program to lower their risk, but the population as a whole would be left alone because the risk of most individuals is too low. Screening programs are expensive and logistically difficult, but a blood pressure determination at the time of a medical encounter is highly cost effective. Since the interventions are not without acceptability issues, by targeting the intervention one spares most people the inconvenience and resources are not wasted on people who do not need them. This Rose called the preventive medicine strategy.

The high-risk or preventive medicine strategy—that of labeling, for example, a certain population as being hypertensive and then intervening just for that high-risk group—has a number of obvious advantages. It means that one can create an intervention that is specifically appropriate for the individual involved, altering the medication and the dietary recommendations to suit that one person. It also means that the person who has now been labeled as high risk is motivated to comply with the intervention. Physicians are also far more motivated to seek those people out because they perceive them as abnormal, and physicians have been trained in medical school to look for the abnormal and not the normal.

To label certain individuals as abnormal and then give them an intervention that they must comply with for the rest of their lives is to label such persons, in a sociological sense, as “deviant.” In other words, they are different—they have to behave differently. They have to stop smoking and alter their diet, but other family members do not have to. That makes them different, and compliance in a situation in which the individual is deviant is always far worse than a situation in which the person can adapt to societal norms in his or her own family and own culture. So there are severe disadvantages to the high-risk screening strategy, although that is the strategy that is most familiar and comfortable to physicians because it also is the strategy used to identify overt disease.

Both major prevention strategies have drawbacks. The first is motivation. It can be difficult to motivate people to engage in preventive behaviors. People may be oriented to accept treatment when they become ill, but maintaining health is often much more difficult. Education is not the same process as behavior modification, and even individuals who are well educated do not necessarily change their behavior on the basis of what they have been taught.

### A Third Mode: Health Promotion

In the late 1970s, a third approach emerged, that of health promotion. This section will discuss health promotion as a basic strategy as

it pertains to individuals. Health promotion has had a particular success in employer-sponsored, worksite programs, which are discussed in Chapter 19. Informally, health promotion can be described as a pragmatic strategy of raising awareness, motivating voluntary change in behavior in ways that enhance health, changing behavior through appropriate incentives and reinforcements, and changing the environment in ways that support changes in behavior. Officially, there are a plethora of formal definitions (some of which are given in Chapter 19) that have in common these same elements and to varying degrees emphasize individual commitment or healthy public policy. The contribution of health promotion, and what makes it different from disease prevention, is the idea of enhancing the existing level of health, whether for the individual or the community, and not just avoiding adverse health outcomes.

The key concept, which is rarely spelled out, is that people should always seek to raise the level of their health and activity from whatever baseline at which they find themselves to a higher level and that communities and nations should raise their population's health status (as defined by the distribution of health risk, or impairment, or other indicators) to the highest level achievable by collective action. This is sometimes called "health enhancement." For the individual, this means a greater capacity to enjoy life and to play a more vigorous role in society. For the community as a whole, this means higher and more equitable distribution of health, by whatever measure, more energy channeled into productive pursuits, and not incidentally, lower healthcare costs (more to invest in other social goods).

As a result, health promotion means different things to different people. To healthy people, it may mean enhancing the level of health and fitness and participating in vigorous exercise, even athletics. To people concerned with their personal health risk, it may mean conventional screening and disease prevention programs. To the ill, it may mean optimal control and tertiary prevention (the prevention of disability) as a means of enhancing their compromised level of health. To the manager, it may mean increased productivity and less absence.

Health promotion is a strategy that in many ways is derived from both the clinical prevention strategy and the public health strategy. It builds on interest in fitness and physical development, on interest in nutrition, and on health education. A commitment to social change is particularly evident in the model as it has developed in Canada, in Europe, and through the World Health Organization (WHO). The concept of health promotion as outlined in the WHO's Ottawa Charter, for example, is to create an environment of choices in which the individual is motivated to pick the healthy choice rather than the least expensive or the most expedient choices in their lives. In the United States, health promotion is somewhat more directive and prescriptive in its approach, and public policy in support of health promotion relies heavily on individual goals set by the U.S. Surgeon General.

The strategy of health promotion is based on the idea that individual interventions can be achieved by motivating health-conscious behavior in subjects as groups and by institutional interventions that change the options available for individual behavior. These interventions may result in an enhanced state of health and well-being and therefore go beyond disease prevention, which seeks only to maintain the current disease-free status. These interventions are not uniquely within the practice of physicians nor are they within the traditional responsibilities of public health professionals. Rather, they exist separately from either the medical model or the public health model. More sophisticated approaches to health education, peer pressure within identity groups such as large companies, and a public rewards system for activities such as marathon running create a social climate in which healthful behavior is not only considered responsible but a social norm. Participation in these normative behaviors is facilitated by incentives and environmental support, which may include low-cost recreational facilities, legal sanctions such as antismoking ordinances, and the introduction of new products by manufacturers responding to a growing market for such items as fitness equipment or preferred foods.

The strategy of health promotion is in part a hybrid of the public health and the clinical preventive medicine approaches in strategy but also a more positive, health-centered approach in its philosophy. In health promotion programs, one uses broad and relatively unselective interventions such as health education, the media, and group activities such as fitness programs to motivate individuals to change their personal health-related behaviors. Decisions are left to the individual with respect to the specific health practices, but programs create persuasive forces to make compliance a social expectation. The result is an approach that allows individuals to make their own, presumably relatively educated, decisions on healthful living and that, at least in the occupational and healthcare setting (less so in the community setting), allows for individual differences.

The result of these changes is an exceedingly powerful and self-perpetuating movement that reverses the traditional context of prevention-oriented services. Before the notion of health promotion was introduced, choosing the more healthful behavior required extra effort and may have been contrary to social norms. Health promotion provides a framework (as advocated by the WHO's Ottawa Charter and its approach to health promotion) to make healthy behavior the accepted standard, so that the unhealthful behavior becomes socially deviant. The health promotion enterprise has set into motion powerful social forces. Concepts about health, physical fitness, and personal responsibility for one's health have been fused with popular culture, so that the message is continually reinforced by peer pressure, advertising, and the approval of the medical community.

Health promotion has the advantages of the public health approach—changes applied to a large population in a behaviorally appropriate context—with the addition of enhanced motivation of the subject and a high benefit-risk ratio in common with the clinical preventive medicine approach but at a lower cost. However, it is not perfect by any means. By making the decision as to whether and how to participate in health promotion activities an individual one, the health promotion strategy falls short in three important ways: (1) it

fails to take into account individual differences that may be unrecognized by the participant, since the ability to benefit from exercise is at least in part genetically determined, (2) it places great and at times unfair peer pressure on those who do not choose to participate, and (3) it functions separately from the healthcare system.

The correct advice to give is not always clear. For example, in the Honolulu heart study, it appeared clear that a reduction in fat intake was associated with a reduction in coronary heart disease. This provided considerable encouragement as a way of reducing individual risk of myocardial infarction. Unfortunately, an evaluation of the same data showed that reducing fat intake was associated with an increased risk of cancer. When the two were put together, it appeared that the population derived hardly any benefit from reducing fat because the one seemed to outweigh the other. Years later, it is recognized that the type of fat intake and the type of cholesterol are important in the effect, and much more is known. At the time, however, the health message was very confusing. This conundrum of uncertainty has recurred many times over the decades with respect to cardiovascular risk factors and remains an issue with respect to moderate levels of overweight.

## MORTALITY AND DISABILITY

Occupational and environmental medicine is not uniquely concerned with mortality. Occupational medicine places more emphasis in practice on reducing the risk of disability. However, concerns about mortality, life expectancy, and disability are important in shaping the goals and strategies of prevention science and so will be summarized here.

### Health and Death

Concepts of age, health, and mortality have changed among the public as well as within the medical community over the last century. Historically, the major preoccupation of medicine was saving lives.

Now, patients are living longer but expect to be kept healthy for longer than they could have reasonably expected in the past. Maurice Chevalier once said, “Old age isn’t so bad when you consider the alternative.” However, growing old, sick, and disabled is not a particularly good trade-off either, but this result has been one consequence of reducing the death rate.

Mortality control (in the sense of reducing age-specific death rates and displacing death into older age groups) has been the great triumph of medicine and public health over the last one hundred years. Many more of us are growing old than in past years.

In 1900 survival declined in a roughly linear curve from young to old adulthood. By 1980, mortality was compressed into increasingly older age groups, and this has continued since. In 1900 one could expect that life might end with a more or less constant probability after surviving infancy until one reached old age. By 1980 by far the majority of people were surviving into their sixties and seventies, but survival was slowing and life expectancy increased more slowly up to today.

The National Center for Health Statistics reports that in 2006, the average life expectancy at birth was 78.1 years in the United States, although with considerable disparity (76 and 70 years for white and African American males, and 81 and 77 years for white and African American females, respectively), and at age 20 (approximately the age of entry into the workforce) it was 58.4 years remaining (56, 61, 51, and 58, respectively, by race and sex). Although it is encouraging that life expectancy is increasing, the United States still has a life expectancy among the lowest of any developed country (others at this level being Denmark, Portugal, and Ireland) and behind some countries with relatively diverse populations (including France, Canada, and Jordan).

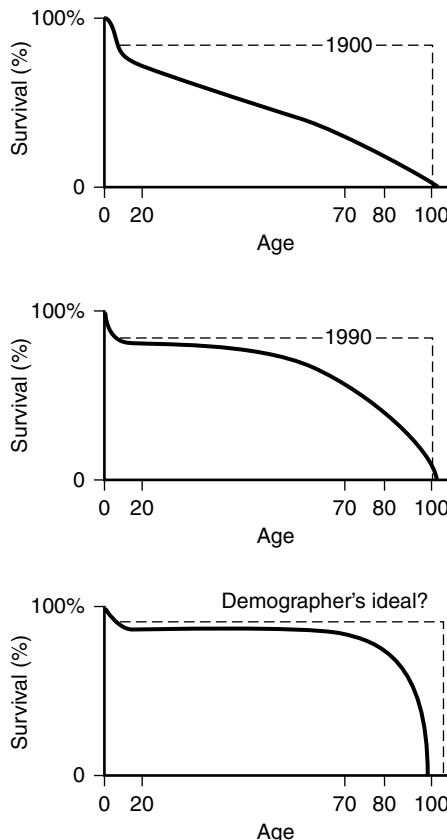
It is a common misperception that heart disease and cancer were rare at the turn of the last century. In fact, heart disease and cancer, and also stroke, trauma, chronic lung disease (usually in the form of bronchitis) and “senility” have always been leading causes of death in the United States. The ten leading causes of death are now heart disease, cancer, stroke, chronic respiratory diseases, trauma, diabetes, Alzheimer’s

disease, influenza and pneumonia, kidney diseases, and septicemia, in that order. Over the twentieth century, rankings of cause of death remained relatively constant until the 1920s, when pneumonia began to fall below heart disease and the infectious causes began dropping out of the top ranks: first diphtheria, then syphilis. In the 1930s, tuberculosis dropped out as a leading cause of death. Chronic lung disease replaced acute lung disease, noticeably after the Second World War.

The net effect of these trends has been to expand survival into the earlier and mid-decades and to concentrate death into a narrower age band at the end of life. Mortality, the frequency of death, is being compressed for the population as a whole into the eighth and ninth decades. This is called, in demographic circles, the rectangularization of the survival curve because the survival curve is relatively steady and then drops off abruptly, like a right angle, close to the natural biological life span (see Figure 8.5).

Ideally perhaps, in the absence of unusual hazards or unforeseen accidents, people would all survive up to a limit that would be biologically determined. There would be no disability or chronic disease or slowly progressing illness. The body would then, in all people, collapse in senescence at the same point in our internally programmed aging process or die after relatively brief illnesses because of frailty and cellular senescence. This idea has often been described as “the one-hoss shay” after an 1858 poem by Oliver Wendell Holmes (the American physician and writer) about a horse-drawn carriage (shay) so finely made that when it reaches the end of its functional life it disintegrates suddenly rather than wearing out part by part. Whether an intrinsic, genetically determined age limit actually exists is an intensely controversial issue in aging research, beyond the scope of this chapter.

This increase in survival and the decrease in mortality at middle ages have not been uniform in past decades or even throughout the developed world. In China and much of the newly successful developing world, life expectancy is already relatively high and is rapidly increasing. Many middle-income countries, such as Argentina and Mexico, now have age-adjusted mortality or life-expectancy rates on a par with countries such as the United States and other members of



**Figure 8.5.** Changes in survival across generations, showing “rectangularization” of the survival curve.

the Organization for Economic Cooperation and Development. Some developing countries have remarkably low mortality rates, including Cuba, Ecuador, and Costa Rica. In Russia, the life expectancy dropped abruptly and tragically, especially for males in the interior of the country, at the time of the breakup of the Soviet Union, apparently because of environmental and social conditions.

Mortality today is confined for the most part to the extremes of age in our present society. Mortality is highest for the very young, particularly at birth and in the immediate postpartum period. It then

falls abruptly to a nadir in late childhood, rises again, steadily increases with a particular bulge for males in their risk-taking youth, and only by old age does it reach the same mortality rates that one observes at birth.

Childhood mortality is dominated by causes associated with congenital and inherited conditions. As a result of the many but individually relatively uncommon inherited problems, congenital defects, and problems associated with the birth process, mortality is initially concentrated in the youngest age groups. These problems are almost all caused by a single underlying defect or agent. In adulthood, by contrast, mortality is increasingly dominated by multifactorial conditions that are much more complex in causality. These multifactorial disorders are also reflected in morbidity. Increasing numbers of adults with these conditions survive them, but many of them are ill or disabled and must cope with a burden of morbidity falling on themselves and on the population.

A well-known observation, often cited in the literature, is that improvement in health status, as measured by such indicators as mortality from tuberculosis, occurred before the introduction of effective treatment. This is often used as an argument (as by Thomas McKeown, using data from Britain) to underscore the importance of social factors, assuming that conventional explanations such as nutrition and improved housing fail to explain the improvement. However, these arguments usually minimize the contribution or conflate medical services and public health interventions and do not respect history. By far the largest absolute and relative drop in tuberculosis deaths in the United Kingdom before the modern era occurred between 1840 and 1850. By this period, Britain was developing a large underclass in the industrial cities, and many people from the relatively healthier rural counties had immigrated to the urban slums. The year 1848 was the beginning of the so-called Sanitary Revolution in the United Kingdom and a decade of great social upheaval and improvement. All through this period, reforms in public health may have been partially effective in reducing the case mortality and illness burden due to tuberculosis: reduced crowding, mandatory air shafts for ventilation in

tenements, case identification and isolation, pasteurization of milk, and reduction in malnutrition. The abrupt drop in deaths from tuberculosis did not occur again until the introduction of effective chemotherapy between 1940 and 1950. During the century between, the trend line is relatively straight. Later, medical interventions (such as induced pneumothorax) that were partially effective did precede the introduction of isoniazid and streptomycin. Obviously, these efforts were not very effective compared to antibiotic chemotherapy, but they did not necessarily have to be as their combined effects took hold and were effective in reducing the mortality from tuberculosis. A similar trend was seen in modern times with the fall in the mortality and lengthening of life expectancy in cystic fibrosis; no one therapy cures the disorder, but incremental improvements have made substantial prognosis. Thus, the case is far from convincing that TB was reduced by some mysterious social factor unrelated to public health or medical services. On the contrary, health or medical services seem to be quite satisfactory as an explanation, although social factors undoubtedly played a role as well.

Living with disability reduces the capacity of the individual for participation in work, family, and society and may bring personal suffering from pain, immobility, and isolation. Disability reduces the quality of life and dis-empowers individuals, depriving them of the opportunity to reach their full potential. The prevention of mortality alone is therefore not enough. Reduction of disability has been recognized as a health priority on an almost equal footing with reduction of mortality. Agencies such as the World Health Organization now track health gains not only in terms of years of life lost or saved, but also in terms of “disability-adjusted life years” (DALYs), applying an adjustment factor that takes into account the quality of life and presence of impairment.

## **THE POPULATION HEALTH FRAMEWORK**

Population health is a broad concept that distinguishes between the health status of an entire population, as measured by various indicators, and the sum total of the health conditions or capacity of the

individuals in the population. It is a Canadian specialty with British roots and has been the organizing framework for unique contributions to thinking about health.

Central to contemporary models of population health has been the health field concept, a way of thinking about the determinants of health that was introduced in a highly influential Canadian federal government document in 1974 that came to be known as the Lalonde Report, after the minister responsible for health at the time. The Lalonde Report viewed health as an issue of inputs and outputs. The health of the individual, and by extension the health status of the population, is determined by four broad “fields” of health determinants: lifestyle, biology (including heredity), healthcare, and environment. The health field concept became the organizing principle for public health agencies in the country for the next two decades.

Population health developed in a framework known widely, especially in Commonwealth countries, as the population health model. The population health model was an invention of a group of investigators, most of whom were associated either with the University of British Columbia or the Canadian Institute for Advanced Research (CIAR). Much of the model is explicitly based on insights from the Whitehall studies conducted by Michael Marmot on British civil servants from the 1970s. The model is outlined in a seminal book on the field, *Why Are Some People Healthy and Others Not?*

The essential elements of the population health model are summarized in Figure 19.1. Building on the health field concept, the model postulates five determinants of health of the population as a whole: genetic endowment (heredity), physical environment, social environment, healthcare, prosperity, and well-being. One of them, genetic endowment, is not amenable to change. Prosperity is much bigger than the health system. The other three health fields were thought to be a sound basis for a public policy that improved the health status of Canadians: health services, environmental influences on health, and personal choice in lifestyle. One could broadly categorize these fields as healthcare, public health, and preventive medicine, respectively.

The population health model also goes into greater depth in considering the social context inherent in the social environment determinant, recognizing such factors as place in the social hierarchy, empowerment (akin to the concept of social capacity in economic development theory), social connections, affluence, and nurturing and early child-rearing.

It is a tenet of the population health model that the social factors on this list, in particular hierarchy and equity in the distribution of wealth but also affluence, are more important as determinants of health for groups of people than any other determinants on the list. The advocates of the population health model infer from the data that there is a nonmaterial social factor that they assign variously to the community, to individual position in the hierarchy, to control over one's workplace or personal life, or to anthropological phenomena, presumably programmed into the primate brain. This draws on the work of Marmot, a pioneering social epidemiologist, and others.

The population health model is a powerful concept in many ways, but it was used to justify a series of reforms that weakened healthcare and tended to marginalize environmental health. The thinking was that economic prosperity would have a general salubrious effect and that health promotion and healthy lifestyle measures would pay off within a single generation, resulting in massive healthcare cost savings.

One criticism of the population health model is that it was put into practice too early, before it was ready. Its advocates had not yet convincingly demonstrated its central tenet, that there is a novel social mechanism driving the attainment of health status in communities. The evidence for the population health model, and for this nonmaterial factor, is based largely on studies that compare indices of health status with income or employment over time in one country or among different countries. Studies based on the population health model suggest that the differences in health status cannot be explained by any of the usual factors associated with income and social class, such as nutrition, lifestyle choices (such as smoking), housing, occupational hazards, access

to medical care, threat of violence, neighborhood pollution, culture and ethnicity, attitudes toward health and disability, exercise and fitness, obesity, and access to daycare and help in child-rearing. These findings require confirmation; the population health model is not empirically validated. However, elements of it, if validated, may lead to innovations and econometric models (relating input of investment in social and economic initiatives to output in improved health status).

From the point of view of OEM, the population health model, at least as developed by CIAR, has many flaws. It seems to underestimate the contribution of occupational and environmental factors to health, possibly because the framers were not thinking in terms of the built environment when they developed it. Furthermore, they did not realize that the environmental influence on health today is as small as it is because people are shielded by effective public health measures from environmental determinants of health, and these public health measures should themselves be counted as important determinants.

## BEHAVIORAL CHANGE

Behavioral change theory can be viewed as a basic science in support of prevention, and as important as epidemiology. General preventive medicine, public health functions involving personal lifestyle, and the critical issue of treatment compliance in clinical medicine all depend on motivating and sustaining behavioral change. Behavior change is critical to success in wellness and health promotion programs (see Chapter 19).

OEM is an exception. For gains to be made in occupational and environmental health, it is important that the workplace or environment change, not the individual. The most effective interventions in occupational health protection—safety science, environmental health, and consumer safety—are precisely those that do not depend on individual behavior: that do not require voluntary action, that are independent of compliance, that cannot be forgotten or ignored, and

that function reliably regardless of the motivation of the actor. That is why “administrative” and behavioral controls rank last in the hierarchy of hazard control, well behind engineering controls (see Chapter 14).

For the OEM physician, a working knowledge of behavioral change is most important for the following applications:

- To design and support programs in wellness and health promotion
- To motivate compliance with those protective measures that do require voluntary action, such as the proper use of personal protective equipment
- To promote safety-related and environmentally sustainable behavior
- To support worker education and training
- To persuade managers and support their decisions
- To motivate compliance in the treatment of occupational injuries and illness
- To communicate with colleagues in other prevention specialties and fields

This section is not a comprehensive or systematic review of the concepts of behavioral change, which is well beyond the scope of this book. Rather, it presents a few selected topics that have been found from experience to be useful to the OEM physician.

### **Health Belief Model**

The Health Belief Model (HBM) is a theory of why people fail to accept a diagnosis, a means of prevention, or a prescription for treatment after they have been educated as to its value and presented with evidence of its effectiveness. The model rests on whether the person considers the goal (i.e., disease prevention) as being of value and whether they believe that the intervention will help them achieve

their personal goals. This is bound up in their perceptions—of their own vulnerability, their own susceptibility, the severity (or dread) of the illness outcome, the obstacles to change, and the benefits of change. These perceptions are conditioned by culture, personality, the psychosocial environment, peer influence, media coverage, and educational and other factors that influence the ability to analyze benefit and risk. However, the most powerful shaping force (and one not well discussed in the literature) is probably the affective response to the perceived threat of the illness or outcome and whether the person is sufficiently scared while retaining confidence of being able to control the outcome. Very high levels of dread might have the perverse effect of paralyzing a person with anxiety and actually promoting the behavior, since many health risk-associated behaviors are stress relieving and subject to the “neurotic paradox” (persistence in a dysfunctional behavior because it reduces stress). The value of the HBM may be more in furnishing a common vocabulary, lists of modifying factors, and a framework for discussion than as a theoretical basis for design of programs. The model was developed by the U.S. Public Health Service in the 1950s.

### **Locus of Control**

The “locus of control” is a general idea from psychology that is often applied to health-related behavior. The concept refers to the belief as to whether one’s life and the good or bad things that happen to a person are under his or her own control (internal) or under the control of others (external). Those with an external locus of control perceive bad things that happen to them or their families to be the result of conditions in the environment, antipathy by neighbors, government policies, the economy, and so forth, and the good things to be, largely, gifts. People with an internal locus of control might see the same bad things as resulting from their personal failings and the good things to be the just results of their talents, skills, and effort. Applied to health, the concept and the instruments based on it are useful in assessing whether a person believes that he or she

can control risks to his or her own health. Those with an internal locus of control are more likely to be compliant with health protection measures, to participate in health promotion activities, and to be proactive in taking care of themselves and their families. Those with an external locus of control are less likely to take personal initiative but may be more likely to take a deterministic approach to their own health risks, assuming that the environment, exposures in the workplace, and (paradoxically, although this aspect has not been much studied) genetics have predetermined their fate. Although locus of control is usually discussed as a general personality orientation, it is quite possible for people to hold different beliefs about different health conditions they may have—understanding, for example, that the development of diabetes may not be under their control, but management of it is. Concepts of locus of control are learned within families and through education, and have a strong cultural component.

The idea of locus of control is closely tied to that of “self-efficacy,” the perceived capacity to act and achieve goals, which in turn is related to how important that goal is in an individual’s life. Locus of control theory is oriented to beliefs about the future, not to explanations of the past (which is called “attributional style”) and should not be confused with personality types such as passive/active or introverted/extroverted. (Even so, it is characteristic for narcissistic personalities to blame external factors for failure and internal factors for success.)

### **Precede/Proceed**

The PRECEDE/PROCEED model is a framework for population-level health education, emphasizing needs assessment and targeted intervention, and is used in the design and evaluation of community-level health promotion programs and social marketing programs. PRECEDE (which originally stood for “predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation”) represents the needs assessment phase, in which one sets goals and

plans for a large-scale intervention. Applied to health promotion, PRECEDE has five steps: (1) problem definition; (2) identification of health determinants; (2) analysis of environmental and behavioral determinants; (4) identification of factors that predispose, drive, or enable these determinants; and (5) designing interventions to modify these factors. PROCEED (for “policy, regulatory, and organizational constructs in educational environmental development”) is the implementation phase. PROCEED adds four subsequent steps: (1) implementation; (2) process evaluation (i.e., how the interventions have been undertaken); (3) intermediate pathway evaluation (i.e., how the interventions have affected the risk factors, with special reference to behavior change); and (4) outcome evaluation (i.e., how the program as a whole changed the frequency of disease, injury, or other outcome). This model was one of many contributions to the field by Lawrence Green and colleagues.

## Social Marketing

Social marketing theory is the application of marketing techniques such as advertising, product placement, persuasion, and media coverage to promote social goods rather than commercial products. It is a major tool for changing behavior in such a way as to promote personal welfare, and it has been extensively used in health promotion. When it has deep penetration in a community, social marketing has the great advantage of creating a new social norm and sense of “social liberation” that allows first individuals and then groups to escape peer pressure and to adopt without disadvantage the new, more healthful or safer behavior. Like commercial marketing, social marketing places great emphasis on researching the market and on keeping the message accessible and fresh.

Social marketing uses mostly the same media tools as commercial marketing but is generally constrained by lack of influence on pricing, packaging (in the few instances in which products are involved), and “point-of-sale” persuasion (intervention at the time the client is making a decision). Even so, some observers find the idea of using

commercial techniques to be manipulative and superficial. Others point out that marketing is not just sales and has a modern tradition of emphasizing relationships and obtaining deep understanding of the “customer’s” needs, using methods of market research, customer feedback on satisfaction and performance, and focus groups. It clearly works, or else for-profit companies would not use it. Why not use it for the good?, the argument goes, especially because all people in the community are subjected to efforts to change behavior in one form or another and from a multiplicity of interests, whether by education, religion, culture, political suasion, or advertising. In that view, social marketing is a positive voice restoring balance and meeting a need.

Social marketing has been heavily and effectively used, for example, in advertising campaigns against smoking and drunk driving. Its first great success came in the summer of 1967, when Sweden switched from driving on the left side of the road to the right. There was the obvious potential for traffic chaos and injury as drivers got used to the new orientation. Supported by a national social marketing campaign simply called “högertrafik” (right-side traffic), the transition went smoothly and motor vehicle accidents actually dropped during the first month, probably because drivers were being careful. Injury rates climbed back later but not much above normal levels, suggesting that drivers relaxed but had successfully learned the new behaviors.

### **Trans-theoretical Model**

The trans-theoretical model defines “stages of change” for an individual with a commitment to changing behavior. As its name suggests, it is descriptive rather than explanatory. The model identifies six discrete stages, based on empirical observation:

1. *Pre-contemplation*, in which the person is not ready to change and is resistant or unmotivated. The person, or client, may be in denial about the level of risk or unconvinced that a change in behavior will be beneficial. People in this stage typically avoid

talking about or thinking deeply about their risks and do not respond to interventions that are based on education. The assumption has been that this group responds best to strategies that “raise awareness”—that is, attempts to “break through” supposed denial (by, for example, showing a diseased lung to a smoker) or to break up apathetic attitudes (by, for example, describing the consequences for the family). A major drawback of the theory is that it does not differentiate among the reasons for people to be in the “pre-contemplation” stage: denial because the idea is psychologically threatening, lack of awareness because the person has not understood the message before, true apathy (which may accompany depression or psychosocial obstacles), distraction because of other priorities in life, demoralization or frustration due to past failures (especially if the person has tried to change the same behavior before), peer pressure to remain the same, fear of being socially “deviant” (unlike other members of the peer group), or cultural acceptance of the risk, among many other reasons. The dynamics are obviously different for each, so “pre-contemplation” means only that the person has not begun the process, and it is not a useful label either for a stage in the process or as a guide to which intervention strategies might work. “Pre-contemplation” is a rather optimistic label because it implies that these people will eventually consider changing, but many never do.

2. *Contemplation*, in which the person is aware of the benefits and costs of change and by definition intends to change within six months. These people are weighing the options and opportunities but are still waiting. This stage is usually thought of in terms of procrastination or ambivalence, and that may be true for most, or even all, people at some point. However, the theory does not address other reasons why these people may be waiting: for more evidence that the benefits outweigh the costs, for an opportunity or triggering event that makes incorporating a new behavior easier, for their spouse to make the same decision, and so forth. Again, because there are different dynamics

for each case and as yet no real commitment to change, classification in this stage is not a reliable guide as to which intervention strategies might be employed.

3. *Preparation*, in which the person has committed to change in principle, anticipates that the person will act within one month. Perhaps a better marker for commitment in principle is the person's taking a concrete step toward action, such as purchasing a self-help book, joining a gym, or asking a physician for help to quit smoking. This is really the first stage that benefits from substantive educational programs and prescriptions for action. Such people do, however, need help in motivation to take the next step.
4. *Action*, in which the person has actually made a change in behavior and has modified his or her lifestyle in a tangible way. Whether the action is effective or sufficient is not considered in the model because the emphasis is on the threshold of demonstrable, as opposed to attitudinal or affective, behavior change. This is probably the stage at which people are most open to education or content-based programs, especially early while the ideas are still fresh and novel and motivation has not flagged.
5. *Maintenance*, in which the person has continued the change in behavior for some period (estimated to last months or years) but is at risk of "relapse" (which in this case means relapsing into previous behaviors and abandoning the change). The primary need here is for the person to have support to continue and to integrate the change so thoroughly in his or her lifestyle that it becomes habitual and no special effort or thought is required to continue. People at this stage need positive reinforcement (such as a reward or social approval) to help them continue. They may need motivation to refresh their commitment, especially if the benefits have not been as dramatic as they expected (a perennial problem with weight loss programs).
6. *Termination*, in which the person has "graduated," does not need further support, and is capable of continuing the new behavior

without external rewards or psychic support. Termination is an ideal, achieved by relatively few people who seek to change their behavior. For the rest, continuation as a lifelong commitment is probably better characterized as periods of maintenance punctuated by periods of real or near relapse during which the person must monitor his or her own behavior, overcome episodes of “temptation” (such as the habit of having a cigarette with coffee or while drinking with friends), and perform tricks of self-discipline. One effective means of self-motivation is to engage in self-administered positive reinforcement, such as by giving oneself a reward, such as a treat or new shoes, for not relapsing.

Table 8.1 describes the types of health promotion programs that are appropriate at the various stages of change. Although not discussed in the health behavior literature, there are many reasons people progress through the stages of change at different rates. One is explained by the theory of “diffusion of innovation.”

### **Diffusion of Innovation**

Diffusion of innovation is sociological theory that describes how innovations and new ideas spread in a community. It can be applied widely—for example, to the adoption of new behaviors, the purchase of technology (such as consumer electronic gadgets), fashion, new business management ideas, and the adoption of new treatments by physicians.

People vary in their acceptance of innovation and new opportunities. Whether from enthusiasm, a higher tolerance for risk, or greater insight because of education or familiarity with the problem, a small proportion of people in any group will adopt an innovation when it first becomes available and will be proponents of change. (The literature calls these people “innovators,” but this is misleading because they did not invent the innovation.) A larger proportion, but still a minority, the “early adopters” will adopt the innovation sooner, gaining experience and reassuring other people that it is worth

**Table 8.1.** Stages of Change in Health Behavior and Appropriate Strategies for Intervention at Each Stage

Stage	Strategy for Changing Behavior*
Pre-contemplation	<p>Objective is to raise awareness and stimulate a person to think about changing behavior. However, because there are many reasons a person may not be considering change, intervention can only be targeted generally. Such interventions usually attempt to break through denial or apathy on the assumption that motivation is the problem.</p> <ul style="list-style-type: none"> <li>• “Consciousness raising”: education, confrontation (e.g., personal testimonials)★</li> <li>• “Dramatic relief”: appeals to emotion (e.g., testimonials of grieving family members, dramatization of struggles with substance abuse)★</li> <li>• “Environmental re-evaluation”: consequences for others (such as family)</li> </ul>
Contemplation	<p>Objective is to move the person toward a personal commitment.</p> <ul style="list-style-type: none"> <li>• “Self-evaluation”: a personal examination of values, beliefs, attitudes, and self-image or cognitive/rational life, with reference to role models (e.g., counseling with therapist)★</li> </ul>
Preparation	<p>Objective is to release the individual from restraints or obstacles to commitment.</p> <ul style="list-style-type: none"> <li>• “Liberation”: adoption of belief that one can change one’s own life★</li> <li>• “Counterconditioning”: substituting a more healthful behavior for the risky one, without significant additional inconvenience, may lead to the stage of Action (e.g., nicotine patch, lean food choices, brisk walks around the block instead of napping)★</li> </ul>
Action	<p>Entire armamentarium of health promotion can be used in the Action phase: stimulus control, contingency management, supportive relationships, counterconditioning, etc., as described above and below.</p>

- “Social liberation”: providing “safe” places and situations and opportunities for social interaction (e.g., clubs and support groups) where behavior can change without opposing peer pressure or openly violating social norms, such as vending machines dispensing condoms or groups that work out together (works especially well for people who are marginalized in some way)\*
- Maintenance
- These strategies are intended to promote adherence to the program after commitment and at least some success. The key is to overcome the perceived cost of continuing to struggle with a change in behavior that imposes some cost (which may be in terms of inconvenience or deviance from social norm).
- “Contingency management”: positive behavioral reinforcement to maintain the behavior, such as encouragement, self-granted small treats, and group recognition (e.g., for weight loss)\*
  - “Stimulus control” to reduce triggers for old, adverse behavior: avoiding or removing cues for the original adverse behavior (e.g., ashtrays)\*
  - “Stimulus control” to add triggers for new, beneficial behavior: putting prompts in the environment that remind one of commitment (e.g., a sign on the refrigerator door)\*
  - “Environmental re-engineering”: changing the environment or use of it in such a way as to support the new behavior (e.g., signs directing to stairs instead of elevator, installing showers in office buildings)
  - “Novelty” and entertainment: to keep person motivated (e.g., magazines and blogs that provide fresh information)
  - “Helping relationships”: support for change in behavior (e.g., family acceptance and encouragement, support groups, social network)\*

---

\* Strategies identified in the work of James O. Prochaska and colleagues.

participating in. The majority will adopt the innovation after it has become familiar and generally accepted. Another minority (“later adopters”) will resist the change but will eventually adopt it. A small number (somewhat pejoratively called the “laggards” in the literature) will be last to or may never adopt the change, whether due to skepticism, a bad experience, distrust, or suspicion. Something similar follows the introduction of a new program in a population, but the phenomenon has not been integrated with the “trans-theoretical” stages of change in the literature.

### **Population-Based Screening Programs**

Screening programs are among the simplest interventions beyond passive health education programs. They are often the first major program that employers adopt. The behavioral challenge is to motivate participation and to maximize penetration (participation rates) in the population at risk.

Employers may conduct screening programs to improve the health of their workers and to demonstrate concern for their health. Hospitals may use screening programs as a “loss leader” to market their institutions. Community-based screening programs have different dynamics than programs based at the worksite (which usually feature high participation rates over a long duration), provided in schools (high participation rates, short duration), and offered by healthcare providers (medium rates, short duration). Participation rates depend on recruitment, and show a progression along the stages of change. Participation is cumulative over time and often starts off slow and gathers speed. It is a multi-stage process, and each stage takes time. Moving candidates into the “Action” stage of participation may require months, and only a minority of people in the community will be prepared to participate in a short-term campaign. Compliance and relapse must be managed and so hospital- or other provider-based programs need a plan, a referral arrangement, or a sustainable infrastructure to ensure lasting benefit.

Literature on participation in screening programs is somewhat misleading because of publication bias and a reliance on single-modality studies. Studies in the literature that claim 80 percent participation by individuals in target groups do not necessarily conform to real life, where 40 percent compliance or participation, for example, in screening programs is doing exceedingly well; however, findings with a lower success rate are difficult to get published. Also, the literature on screening programs is dominated by evaluations of programs devoted to a single screening modality or screening-intervention pair. As the protocol encompasses more interventions, becomes more complicated and potentially confusing, and requires more time and effort from the client, compliance goes down, even if the additional modalities bring much more benefit. Therefore, a screening program that involves five or six modalities will predictably have lower compliance rates than a single program that features, say, mammography.

Newly introduced programs tend to start slow and gain momentum as visibility in the community and word-of-mouth grow, attracting more participants. Penetration rates are usually highest in remote, small, socially cohesive communities, probably because of peer pressure and community interest. The Cochrane Collaboration has identified three interventions that increase participation rates in community-based screening programs: letters of invitation, advance mailing of health educational materials, and telephone prompts.

## RESOURCES

- Rose G, Khaw K-T, Marmot M. *Rose's Strategy of Preventive Medicine*. New York: Oxford University Press; 2008.
- Shumaker SA, Ockene JK, Riekert K, eds. *The Handbook of Health Behavior Change*. 3rd ed. New York: Springer; 2009.
- Wallace R, ed. *Maxcy-Rosenau-Last Public Health and Preventive Medicine*. 15th ed. New York: McGraw-Hill; 2007.

## NOTEWORTHY READINGS

- Epp J. *Achieving Health for All: A Framework for Health Promotion*. Ottawa: Health and Welfare Canada; 1986. [Successor to the Lalonde Report]

## The Praeger Handbook of Occupational and Environmental Medicine

- Evans RG, Barer ML, Marmor TR. *Why Are Some People Healthy and Others Not? Determinants of Health in Populations.* New York: Aldine de Gruyter; 1994.
- Friedman DJ, Starfield B. Models of population health: Their value for US public health practice, policy, and research. *Am J Pub Health.* 2003; 93:366–369.
- Guidotti TL. “Why are some people healthy and others not?” A critique of the population health model. *Ann RCPSC.* 1997;30:203–206.
- Guidotti TL. Health promotion in perspective. *Can J Public Health.* 1989; 80:400–405.
- Gunning-Schepers LJ, Barendregt JJ, van der Maas PJ. Population interventions reassessed. *Lancet.* 1989; i: 479–481.
- North FM, Syme L, Feeney A, Shipley M, Marmot M. Psychosocial work environment and sickness absence among British civil servants: The Whitehall II Study. *Am J Public Health.* 1996; 86:332–340.
- Shah CP. *Public Health and Preventive Medicine in Canada.* Toronto: Saunders Canada; 2003.
- Wall S. Epidemiology for prevention. *Int J Epidemiol.* 1995; 24(4):655–664.

**THE PRAEGER  
HANDBOOK OF  
OCCUPATIONAL AND  
ENVIRONMENTAL  
MEDICINE**

*This page intentionally left blank*

# **THE PRAEGER HANDBOOK OF OCCUPATIONAL AND ENVIRONMENTAL MEDICINE**

---

**VOLUME**

**II**      Central Issues

Tee L. Guidotti, MD, MPH



AN IMPRINT OF ABC-CLIO, LLC  
Santa Barbara, California • Denver, Colorado • Oxford, England

## **Copyright Acknowledgments**

Material from two chapters adapted from Tee L. Guidotti. "Taking the Occupational History." *Ann Intern Med.* 1983; 199-4: 641-651.

Copyright © 2010 by Tee L. Guidotti

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, except for the inclusion of brief quotations in a review, without prior permission in writing from the publisher.

## **Library of Congress Cataloging-in-Publication Data**

Guidotti, Tee L.

The Praeger handbook of occupational and environmental medicine / Tee L. Guidotti.

p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-313-35999-6 (set hard copy : alk. paper)—ISBN 978-0-313-36001-5 (vol 1 hard copy : alk. paper)—ISBN 978-0-313-36003-9 (vol 2 hard copy : alk. paper)—ISBN 978-0-313-38204-8 (vol 3 hard copy : alk. paper)—ISBN 978-0-313-36000-8 (set ebook)—ISBN 978-0-313-36002-2 (vol 1 ebook)—ISBN 978-0-313-36004-6 (vol 2 ebook)—ISBN 978-0-313-38205-5 (vol 3 ebook)

1. Medicine, Industrial—Handbooks, manuals, etc. 2. Environmental health—Handbooks, manuals, etc. 3. Medicine, Industrial—Handbooks, manuals, etc. I. Title. II. Title: Praeger handbook of occupational and environmental medicine.

[DNLM: 1. Occupational Medicine. 2. Environmental Health. 3. Environmental Medicine. 4. Occupational Health. WA 400 G948p 2009]

RC963.32.G85 2010

616.9'803—dc22 2009046237

ISBN: 978-0-313-35999-6

EISBN: 978-0-313-36000-8

14 13 12 11 10 1 2 3 4 5

This book is also available on the World Wide Web as an eBook.

Visit [www.abc-clio.com](http://www.abc-clio.com) for details.

Praeger

An Imprint of ABC-CLIO, LLC

ABC-CLIO, LLC

130 Cremona Drive, P.O. Box 1911

Santa Barbara, California 93116-1911

This book is printed on acid-free paper 

Manufactured in the United States of America

# CONTENTS

## VOLUME 1: PRINCIPLES

<b>Preface</b>	<b>vii</b>
1. Occupational and Environmental Medicine	3
2. Toxicology	55
3. Epidemiology	201
4. Assessment of Exposure	295
5. Monitoring, Surveillance, and Screening	327
6. Safety and Injury Prevention	357
7. Risk Science and Risk Communication Skills	379
8. Prevention Science	459

## VOLUME 2: CENTRAL ISSUES

9. Physical Hazards and Conditions	505
10. Common Chemical Hazards	569
11. Buildings	689
12. Environmental Health	739
13. Stress and Psychosocial Factors at Work	809
14. Hazard Control	845
15. Clinical Evaluation	889
16. Occupational Injuries and Musculoskeletal Disorders	949

## VOLUME 3: PRACTICE INSIGHTS

17. Diseases and Health Protection	1013
18. Capacity for Work	1131

## **Contents**

19. Wellness and Health Promotion	1201
20. Occupational Health Services	1245
21. Workers' Compensation	1315
22. Emergency Management	1357
23. Medicolegal Services	1389
24. Regulatory Agencies	1443
25. Society, Critical Science, and Ethics	1481
26. Global Occupational and Environmental Health	1537
<b>Conclusion</b>	<b>1585</b>
<b>The Essential Library for the Occupational and Environmental Physician</b>	<b>1597</b>
<b>Index</b>	<b>1599</b>

# **VOLUME 2:**

## **Central Issues**

*This page intentionally left blank*

# **9 PHYSICAL HAZARDS AND CONDITIONS**

Physical hazards have in common the storage and release of energy in various forms: mechanical energy, sound, vibration, electromagnetic radiation, heat, pressure, and so forth. The storage of energy is “potential energy,” such as a ball held over the floor. When the ball is released, the potential energy becomes “kinetic energy” due to the pull of gravity. When it hits the floor, the ball is temporarily deformed and energy is stored again. When it bounces up, that potential energy is released as kinetic energy again.

Physical hazards are concerned with the uncontrolled release of energy, and the harm or injury that kinetic energy can do. This release may be sudden and uncontrolled, as in an explosion or mechanical safety hazard, or sustained and controlled, as may occur in conditions of work that expose the worker to noise or heat. The sudden, uncontrolled release of energy usually involves the physician only insofar as the treatment of trauma is required. In the division of labor in occupational health and safety, that domain is the responsibility of safety engineers, and physicians rarely play a role.

Kinetic energy includes electromagnetic energy, electricity, heat, mechanical movement (such as rotation), sound, and wind pressure. Potential energy includes magnetism, gravity, chemical energy,

nuclear energy, and mass (which is a known equivalency, thanks to Einstein).

This chapter will address the health implications of various physical conditions of work, with an emphasis on long-term exposure to various physical hazards.

## ERGONOMICS

Ergonomics can be defined as the scientific discipline and practical field of designing the immediate environment to adapt to the performance characteristics of human beings with the objective in occupational health of matching the workplace to the capabilities of the worker. The goals of ergonomic design are safety, efficiency, comfort (which implies sustainable energy expenditure), productivity, ease of use (emphasizing intuitive functionality over the need for complicated instructions), sensory and perceptual efficiency (signals easily perceived and understood), and aesthetics (which is not trivial because a tool or device a person does not like will not be freely used). Optimizing these goals is very challenging and at its most sophisticated ergonomics blends seamlessly with art, architecture and design, and engineering. The domains of ergonomics are generally described as physical (pertaining to biomechanical factors, safety, materials handling, and posture), cognitive (pertaining to perception, mental processing, response time, and motor function), and organizational (pertaining to work design and organizational structure). The reader seeking more information is encouraged to consult a textbook on ergonomics.

This chapter will limit its treatment to physical ergonomics. Physical hazards and mechanical forces represent only one dimension of ergonomics, but an important one that is most closely related to the occupational and environmental medicine (OEM) physician's practice. Physical ergonomics is a branch of knowledge involving the interaction between people and the artificial physical environment. Ergonomics is briefly described in historical terms in Chapter 6, with an emphasis on the human factors movement. In its modern incarnation, it has advanced beyond the more limited tradition of human

factors research and has established itself as an occupational health profession in its own right, specializing in the design of the workplace and workplace procedures for optimal comfort, productivity, safety, and the minimization of human error.

Physicians generally do not practice ergonomics. Ergonomic problem-solving is generally done by professional ergonomists, safety specialists, or another individual, often informally trained, who works with the guidance of one of the many handbooks and Web sites available to make a workplace easier to use. The OEM physician needs to know enough to be able to make recommendations with respect to simple problems and know when to advise referral to a professional ergonomist. A grasp of ergonomic principles is also important to understanding repetitive strain injury and musculoskeletal disorders, which are discussed in Chapter 16.

The major insight of ergonomics is that people and the artificial environment “fit” together in ways that are much more complex than biomechanical dimensions. When a worker sits at a workstation with monitors and controls, uses a tool, performs a task for a prolonged period, or concentrates on a message or received information, a complex interaction is taking place that is only partly described by the placement of the worker and the energy being expended. The reach to a control may be convenient, awkward, or a severe strain. A tool may be poorly designed to fit in the hand and fatiguing to use. The task may be poorly designed and may place excessive stress on a weak part of the body. The incoming information may be difficult to decipher immediately because it may lack clues to meaning or emphasis, or may be placed in a distracting context. In each case, attention to the design of the worker-environment interface may reduce fatigue, increase productivity, and reduce the likelihood of error.

With changes in employment policies and the drive for productivity, assembly and other repetitive tasks are frequently “de-skilled,” or broken up into simpler standardized tasks that are easy to learn and supervise. Employees are often “multitasked,” or cross-trained in several functions. The result may be that they are treated by management as interchangeable, standardized parts themselves. This may result in

inflexible workplace design that reduces efficiency, and may also result in boredom, alienation, and loss of job satisfaction. The outcome may be a paradoxical fall in productivity. A modicum of control over the immediate environment is a powerful motivating factor among workers and restores efficiency, in addition to making the work more rewarding.

Worker characteristics are critically important in ergonomics. In the past, most workplaces were designed by engineers, managers, or architects who had little or no applied knowledge of ergonomics or anthropometrics. As a result, workplaces tended to be appropriate for young, fit men of average to tall stature, but were often awkward and uncomfortable for the worker actually hired into the job. The capacity of workers also changes with age: reduced visual and auditory acuity, increased time for processing and reacting to information, and reduced strength make jobs designed for young workers difficult for older workers.

Workplace design is a major part of ergonomic practice. Workplaces are typically designed to accommodate machines and to facilitate supervision. Often, management attempts to hire workers who fit the job or to overcome design problems with increased training. A more productive approach is to design a workplace that can be modified and adapted to the needs of individual workers. Whenever possible, it is preferable to change the environment rather than the characteristics of the worker. Tasks requiring upper body strength to pull levers may be redesigned for women by using foot pedals, for example. The interventions will be more reliable and will protect a greater number of workers.

Another insight from ergonomics is that good design has more far-reaching effects than simply to overcome obstacles and accommodate limitations. Tools and workplaces designed for older workers and workers with impairments are almost always easier, safer, and more efficient for younger and fit workers as well. If the workplace is designed for workers with modest limitations, there will usually be more inherent flexibility and safety in the job and a greater potential for substituting workers of different ages and backgrounds. Sound ergonomic design incorporates conventional control of safety hazards,

but also reduces the potential for fatigue, miscalculation, distraction, and cumulative trauma.

The workplace should be under the control of the worker to the extent necessary to provide adequate illumination, comfortable seating, easy clearance for turning and bringing work in and out, and a comfortable climate. Experienced ergonomists have numerous “tricks of the trade” for improving workplace design. These include numerous computer adaptations (adjustments and peripherals, such as keyboards and mice), footrests (which make prolonged sitting much more comfortable), back supports, lighting, stools (making it easy to sit and stand in quick succession), and means of organizing work in bins and trays. These modifications are designed to be simple, accessible, and low in cost because otherwise they will not be used; “high-tech” solutions are generally undesirable in ergonomics because they are costly, obsolesce quickly, and often require maintenance and updating.

In offices and other non-production workstations, the key to successful intervention is sustainability of static activity, designing the placement of the work surface, reaches, seat, and visual displays in a way that is compatible with the measurements of the worker’s body so that as many actions as possible are in neutral position and require minimum energy. In situations where modification to a particular individual is not practical, the usual approach is to design for a range of anthropometric dimensions that accommodate the great majority of workers.

On the other hand, the key to successful intervention in production workstations is sustainability of dynamic activity so that movement is efficient, requires manageable energy expenditure, and does not strain the capacity of the worker. As mentioned in Chapter 6, at the turn of the twentieth century there was a strong movement among employers to design workplaces and to define tasks for efficiency, based on the work of Taylor. This was interpreted at the time, and to a large extent, still is, as design conducive to short-term efficiency and maximal production. The problem soon arose that a workplace designed for maximum production could be unsustainable over time. Workers’ bodies could not keep up, nor could workers

sustain focused attention on repetitive tasks without error for as long as was expected of them. During the years of World War II, especially, many of these ideas were reevaluated because of human factors research in the military, but were later forgotten or discounted. With the advent of information technology, a new era of tasks requiring repetitive movements and workplaces designed for efficiency re-created many of these problems in a new, high-tech environment.

Workplaces should be designed for comfort and reduction of fatigue, not for efficiency alone. Standards of illumination, vibration, temperature, and humidity, which are applied by architects and builders, are intended for the population mean and are frequently outside the comfort zone of some individuals. The strain, stress, and distraction that result are expressed as fatigue, eyestrain, or vague discomfort. Typically, these complaints affect only a small minority of workers and tend to be exaggerated when there is psychological stress and dissatisfaction. Discomfort, and the inevitable conclusion that management does not care, then feed into the adverse psychosocial conditions of work, discussed in Chapter 13.

Personal interventions can also be made. These interventions can be physical. Footgear that resists slippage on slick flooring provides better footing, reduces the likelihood of falls, and allows more to be carried without fatigue. As noted, ergonomics merges safety science and injury prevention behavior in the safe design of workplaces.

Proper training of supervisors and workers is the most important and flexible approach to the solution for the worker. Training in the performance of well-designed tasks using suitable tools in an appropriately designed workplace is very effective in increasing productivity and reducing errors. Frequent rest stops may result in much greater productivity than sustained production, with fewer injuries and less fatigue. Training alone cannot overcome serious workplace design flaws, however.

Tasks and tools can also be designed for optimal efficiency and productivity. Unusual, awkward, and repeated motions should be minimized or avoided. Workers who are left-handed should have left-handed tools available (notably scissors), especially if they are used

frequently or require force. Certain types of tools may require constant energy expenditure by the worker over prolonged periods, such as wrenches or forceps. In general, the less static contraction of muscles, the longer contraction can be maintained. Poorly designed tools can often be replaced by spring-loaded, leveraged, or locking devices that are less fatiguing to use. Tools that fit the hand poorly or require awkward positions to use can usually be redesigned or replaced by more biomechanically efficient models from commercial sources. Controls can even be shaped so that their tactile qualities can be felt subliminally, further reducing the likelihood of error.

Aesthetics and comfort matter a great deal in the workplace. A tool that feels awkward in the hand, such as a pen or stylus that is too small or a knife with a poor grip, will be avoided by the worker. Nor does anyone like to work in an ugly place, in part because it conveys a message that management does not care. Aesthetics can be overemphasized, of course, and there are numerous examples of places designed primarily for beauty that turned out to be nonfunctional. (The masterpiece houses designed by Frank Lloyd Wright, for example, are all high-maintenance and difficult to live in.) A misplaced devotion to the “quality of working life” may also be used as a substitute for good workplace design (see Chapter 25). Aesthetics, and even artistry, are qualities to be admired in ergonomics, but in the workplace aesthetics should serve functionality, not compromise it.

Information processing is made much easier and more accurate when presented to the senses in a way that facilitates rapid perception and cognitive interpretation. Instruments can be clustered and aligned such that deviations from the expected are immediately obvious. The selection of type fonts, colors, coding schemes, visual cues, and labels is an inexpensive but highly effective way to increase efficiency and reduce errors in performing complex tasks. Likewise, visual displays can be designed so that unusual or urgent information is easily and rapidly interpreted by the use of color-coding or analog displays, rather than as a continuous reading on a meter. In this application, aesthetics can be used in reverse. An alarm, for example, should be ugly, perhaps using clashing colors or an irritating tone, not soothing or subtle.

An important application of ergonomics is in rehabilitation and workplace modification. The application of ergonomic principles makes it possible, usually at low cost, to accommodate the needs of disabled or recuperating workers and to ensure their smooth transition into working life, sustainable productivity at work, and even employability. This issue is discussed more fully in Chapter 18.

Practical ergonomics is beyond the scope of this book, but there are some useful basic principles that the OEM physician can readily apply to the most common inquiries, which are about office workstations:

- Adjust the workstation before starting to work; the time to improve the office ergonomically is before, not after, symptoms appear. Nonadjustable working surfaces are far too common but can usually be modified at least to some degree; one might try working for a half hour to identify problems and then see how the workstation can be changed. For example, people of short (otherwise known as “ideal”) stature may find it more convenient to raise the chair and to use a footrest than to lower the keyboard level. Telephone books can raise monitors to a better height. Such ad hoc arrangements sometimes shame managers into getting better equipment.
- The most sustainable position for continuous or repetitive work is that closest to “neutral,” which is the natural, resting position of the body part. The more deviation from neutral position, the more energy is required to sustain the posture or position, and the more tiring and unsustainable it is.
- For sitting in front of a keyboard, neutral position is arms by the sides and elbows at 90°, wrists minimally bent. This absolutely requires that the keyboard be positioned lower than the standard desktop.
- Awkward or unnatural body movements, such as bending and twisting at the same time, or reaching behind the back, should be eliminated by reorganizing the work space.

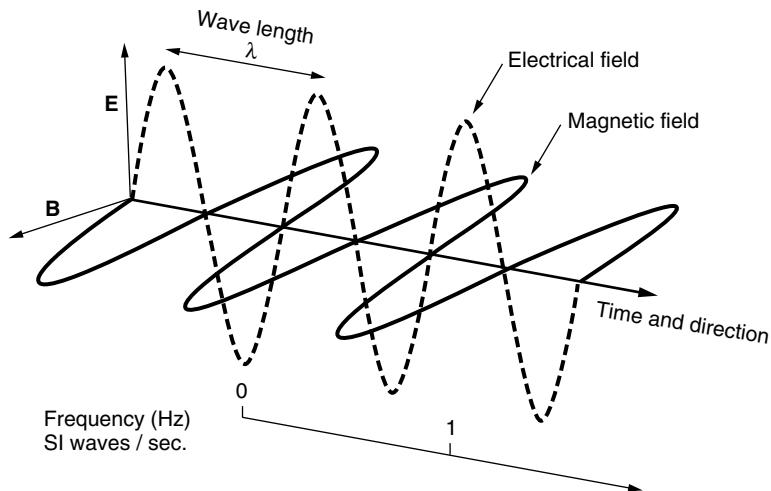
- Repeated exertion of an intrinsically weak body part (such as the fingers or wrist) risks straining the soft tissues and causing repetitive strain injury.
- Sitting posture needs to be assisted with adjustable chairs such that the back can be kept straight and supported, and the worker does not have to bend over to do the work.
- Armrests should be adjustable or removable if possible, never too high; this causes upward pressure on the shoulder and is uncomfortable when the worker sits back.
- The seat pan of chairs should be low enough that the upper thigh does not hang on the edge. The seat should be deep enough that the worker is not sitting forward without back support, but not so deep that the edge of the seat compresses the back of the thigh or the area behind the knee. There should be enough space to insert a fist between the knee and the edge of the seat while sitting comfortably.
- The edge of the seat should be rounded or padded, not sharp or squared off.
- Feet should rest flat on the floor. If they do not, a footrest, which can be as simple as a sturdy box, is required.
- Chairs should have a swivel base with five casters; a chair with four casters is unstable and can tip over easily.
- Many short breaks are better than a few long breaks; ideally, about 10 seconds every five minutes should involve a motion other than using the keyboard. Longer breaks of several minutes should be taken every half hour to an hour.
- There is no tracking instrument that suits everyone; the most suitable is the one the worker finds most comfortable. The mouse should be used in neutral position, right beside the keyboard, with no reaching. Track balls are particularly good alternatives for people with impaired finger motion.
- Keyboards are designed to fit most, not all, people's hands. If a standard keyboard does not work for an individual, there are a

variety of ergonomic keyboards available; the most suitable is the one the worker finds most comfortable.

- Monitors should be squarely in front of the worker's face, in the natural line of sight with the head in neutral position. If it is too low, the worker will bend over slightly or hunch; if it is too high, the worker will bend his or her neck backward and may experience neck pain. If a workstation is going to be used by many workers, consider obtaining a monitoring arm that mounts to the wall or clamps to the desk, suspending the screen so that it can be adjusted easily.
- Document holders should be next to the keyboard and well lit. It should not be necessary to lean over in order to read while inputting data or text.
- Laptops are more difficult than desktop computers to use for prolonged periods. Consider a docking setup when in the office, so that the laptop can be plugged into a full-sized keyboard, mouse, and large screen.
- Lighting should be adjusted so that it is easy to read words on paper but there is no glare reflected off the computer screen. Desk lamps can be equipped with dimmer switches.
- Glare can be minimized by positioning light above and behind or to the side of the worker, and by using indirect or shaded light. Avoid placing a monitor in front of a window where there is bright daylight or where a window or other light source is directly behind the screen. Glare can also be controlled by anti-glare film screens applied to the monitor, by fabric screens, and by curtains over windows.

## THE ELECTROMAGNETIC SPECTRUM AND MAGNETIC FIELDS

Electrical fields (conventionally called **E**) and magnetic fields (conventionally called **B**) are linked: they propagate together and perpendicular with one another along the axis of their transmission



**Figure 9.1.** Electrical (**E**) and magnetic (**B**) fields are integrally linked and perpendicular to one another.

(conventionally called  $Z$ ). Figure 9.1 illustrates the relationship. As with sound or any other wave, they can be characterized by frequency ( $\nu$  in Hz, which has units of  $s^{-1}$  and is known as “cycles per second”), wavelength ( $\lambda$ , in m), and energy delivered. The relationship between wavelength and frequency is fixed and inverse; the product is a constant (the velocity at which the wave is propagating in the medium,  $v$ ). Very simply:

$$v \times \lambda = \nu$$

The usual metric measure of wavelength varies, for convenience, with the place on the spectrum: meters (radiofrequency waves), centimeters (most common for microwave, because 1 cm is the optimal wavelength for heating water), and micrometers or nanometers (for infrared, visible light, and ultraviolet).

More energy is carried in the shorter wavelengths, corresponding to higher frequencies. Figure 9.2 illustrates the electromagnetic spectrum and its subparts. Electromagnetic radiation above 500 nm (in

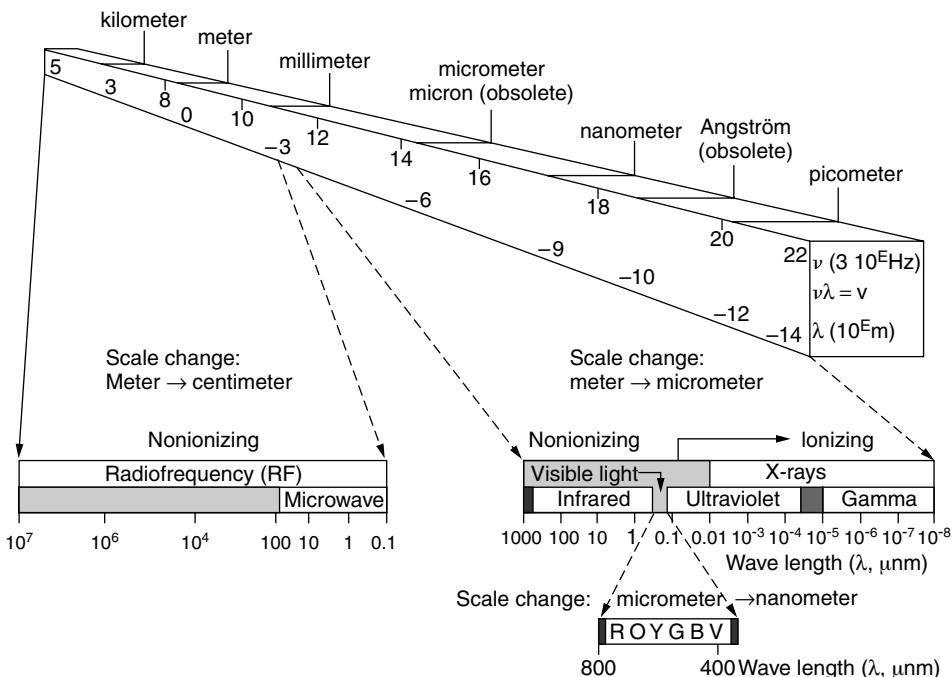


Figure 9.2. The electromagnetic spectrum.

the ultraviolet part of the spectrum) carries sufficient energy (greater or equal to about one electron volt) to detach tightly-bound electrons from atoms, creating an ion. The higher the energy level, the greater the ionizing effect. In water, and therefore in biological tissues, these ions can cause intracellular damage; and, when the energy level is high enough, cascades of free radicals can be formed. Electromagnetic radiation below that level is sufficient to impart energy to, or to "excite" atoms, but not to ionize them.

A large part of the electromagnetic spectrum is of little interest in OEM because of the absence of documented biological effects. These frequencies include the so-called extremely high frequencies (EHF, 30–300 GHz, wavelength in millimeters), super high frequency (3–30 GHz, wavelengths in centimeters), and ultra high frequency (0.3–3 GHz, wavelength in tenths of a meter). UHF is familiar as an

ancillary spectrum previously used for television transmission; “very high frequency” (VHF, 30–300 MHz) is the standard television transmission range and lies just under the microwave channels. These frequencies are not known to have biological effects, although there has been speculation.

Magnetism is a fundamental force in physics. Static magnetic fields are generated by dipole magnets, such as the Earth itself. Static magnetic fields induce electrical current in conducting materials that pass through them, in a direction perpendicular to the movement across field strength lines (see Figure 9.1). Dynamic magnetic fields are generated by current-carrying electrical conductors.

Magnetic field strength, which is measured in gauss (G), may be quite independent of electrical fields; both tend to fall off rapidly with distance from the source. Magnetic fields of strengths greater than the background field of the Earth are ubiquitous in urban life. Sources include building wiring, appliances, power distribution lines, and ground currents in steel beams, as well as plumbing and gas lines universal in industrial settings and homes. Until recently, no health effects were thought to be associated with exposure to magnetic fields, at least at intensities encountered in industry and the natural environment.

In 1979, a pilot study by Wertheimer and Leeper suggested that children who had died of cancer in Denver were more likely to have lived in houses that had higher current levels from nearby distribution lines, which generate higher local magnetic fields. Subsequent efforts to refine the methods and obtain additional data strengthened their findings, and a small number of additional studies have also reported an association; however, most studies, including some massive epidemiological studies, have not. The effect, if one exists, seems to be strongest for childhood leukemia. Adult cancers have been more extensively investigated in studies of electrical workers and radio transmission operators, and have shown few consistent indications of an excess of any particular cause of death. The studies have suggested but not confirmed an association with brain cancer and, more weakly, an excess of adult leukemia.

A major problem with this line of investigation is accounting for the relative weakness of the fields and the presence of other, confounding exposures. Laboratory investigations have not yet demonstrated a plausible biological mechanism for effects from magnetic fields, but have suggested alterations in cell membrane characteristics that might have some bearing on cell function. At present, health implications of exposure to magnetic fields remain speculative.

## **IONIZING RADIATION**

Nonionizing radiation is discussed in a separate section in this chapter. This section addresses occupational health concerns specifically associated with ionizing radiation. Ionizing radiation is also a significant issue in environmental medicine.

The primary occupational risk of radiation-induced health effects is encountered in mining, the nuclear reactor industry, the nuclear weapons industry, industrial settings in which shielded radiation sources are used (for example, to check the integrity of pipeline welds), and medical occupations involving diagnostic x-ray or radiation therapy. Radiation health, also called “health physics,” is a specialized area of occupational health and safety with its own qualifications and educational programs. However, radiation hazards, especially at low levels, are common enough that the OEM physician should have a general familiarity with the health implications of exposure to ionizing radiation.

Ionizing radiation carries sufficient energy to dislodge electrons on impact from their otherwise stable orbital cloud. This results in the production of charged ions or free radicals, both of which are highly reactive and therefore toxic on a cellular level. The process also produces secondary x-rays as well as displaced electron showers. These are relatively weak effects not known to be biologically significant, but the principles behind them apply to scanning electron microscopy, micro beam analysis, some occupational hygiene methods, and other measurement and imaging techniques of importance in OEM.

Ionizing radiation occurs in two forms: electromagnetic radiation and particle radiation.

- Electromagnetic radiation is a beam of energy carried by photons, observed in the form of x-ray or gamma ( $\gamma$ ) radiation of about 100 nm or shorter.
- Particle radiation consists of atomic fragments or particles, observed in the form of alpha ( $\alpha$ ) or beta ( $\beta$ ,  $\bar{e}$ ) radiation, or neutrons ( $n$ ).
- Alpha particles consist of two protons and two neutrons (the stripped nucleus of a helium atom) and are heavy, slow, and poorly penetrating, although they carry much energy.
- Beta particles are stripped electrons or, rarely, positrons ( $\beta^+$ ), and are fast and light but poorly penetrating. As they pass through lead and other dense materials, they produce x-rays as they lose energy and interact with matter.
- Neutrons are relatively heavy, carry much energy, and are deeply penetrating. They are a practical concern in nuclear fission only, and are not produced by nonnuclear processes.

### **Sources and Exposures**

Sources of radiation on Earth include natural radionuclides (isotopes in which spontaneous decay results in emission of a particle and x-ray), outer space (cosmic radiation), medical and industrial x-ray technology, fallout from above-ground nuclear weapons testing, and nuclear fission and fusion. Approximately 82 percent of the radiation dose received by the U.S. population is of natural origin, over half of which results from the decay of radon; the remainder comes in roughly equal amounts from cosmic radiation, terrestrial sources in rocks and soil, and internal sources in the body. Anthropogenic sources are found mainly in medical applications.

Occupational exposures are concentrated in a relatively few but expanding number of industries, including healthcare, energy

production, mining, metal products and fabrication, and defense. Cigarette smoke adds exposure to radionuclides (primarily polonium) in quantities that are small but densely concentrated on the bronchial epithelium. The hazards of the nuclear energy industry and of research and technological development in fusion and high-energy physics are a highly specialized area of concern usually handled by health physicists and specially qualified radiation health officers. A detailed discussion of these industries is beyond the scope of this book. Occupational physicians working in the industry should receive special training.

Exposure to radiation may be “whole body” or may be limited to a particular body part. Whole body radiation might occur because of an exposure to an unshielded source in an uncontrolled incident. Limited or segmental radiation may come from an external source when a worker passes in front of the beam or if the shielding is faulty. Limited radiation may also arise from an internal source such as inhaled radionuclides that may either remain in the lung or migrate to bone. Alpha emitters that accumulate in lung or bone, such as plutonium (Pu) and radon (Rn) daughters, are of particular concern in this regard because they emit radiation of high energy that, while not penetrating far, intensively interacts with vulnerable cells in the unprotected tissue. For this reason there is a particular risk of lung cancer or bone cancers from alpha emitters taken into the body from activities such as uranium mining and nuclear weapons production accidents.

A key concept in the understanding of health effects from radiation is “linear energy transfer” (LET), the average energy deposited per unit length of the track of the particle or photon as it passes through matter (most of which is empty space at the molecular level) and interacts with one molecule and then another. This concept can be understood by analogy to shooting a ball on a pool table with great force: the heavier the ball and the faster the ball is set in motion, the more energy it imparts to each ball that it hits. The scattering of these balls then imparts energy to other balls. The LET is used in calculating the “relative biological effectiveness” (RBE) of a radiation

beam because it determines the ionizing potential of the radiation in cellular water. In a cell, the energy released results in cascades of ions or free radicals, which exert the biological effect. The magnitude of damage caused by the passage of radiation through tissue is closely related to the LET. “Low-LET” radiation, such as x-rays, causes little injury; “high-LET” radiation, such as neutrons or alpha particles, causes much. The production of ions, free radicals, and free electrons from water in the cell results in an “indirect effect” that ultimately may damage a critical molecule in the cell. The interaction of radiation itself and a critical macromolecule such as DNA is a relatively improbable event and results in a “direct effect.”

### **Measurement and Assessment**

Radiation is measured in many ways. In order to understand the various measurements, it is important to go back to first principles and to define what they mean. However, the system of measurement has recently passed through an extensive revision in which several of the units used in the past have been converted to units in the Système internationale (SI), the international system of scientific units, of which the metric system is a part. Thus, the literature of OEM and many regulations (particularly those of OSHA) are written in the old units; but the OEM physician must be conversant with the new measurements if he or she works in any sector involving radiation. The current generation of OEM physicians will simply have to live with two coexisting systems.

In occupational medicine, the most important measurements are or have been the roentgen (R); the rad, which is now being replaced by the gray (Gy); the rem, which has been replaced by the sievert (Sv); and the working level (WL), which is used in exposure regulation and occupational epidemiology. For reference, a medical x-ray exposes the patient to about 0.4 mSv, or 39 mrem, of radiation.

Measurement of radiation takes into account four sequential steps: emissions from the source, exposure of the object, absorption by the object, and the dose of radiation that results. Each is measured

differently and using different units. The new, SI units do relate to one another. The old system was not as systematic.

Measurements of activity, or emissions from the source, quantify the strength of the radiation source in terms of emission of radioactivity per second. The strength of a source of radiation is properly measured in SI units of the becquerel (Bq), which is the measure used in Department of Energy regulations and reflects the culture of physics and engineering. One Bq is one event of radioactive decay per second. This is a very small unit, because any radioactive source in the real world will consist of billions or trillions of atoms. OSHA regulations, however, use units of curies (Ci), which are much larger (by nine orders of magnitude) and are being phased out by scientists. One Ci equals  $37.04 \text{ GBq}$ . The main advantage of using Ci is that it is scaled more appropriately for human experience.

Measurements of exposure quantify what reaches the object (or person) being exposed and is expressed as a flux, in terms of the amount of radiation that produces a given electrostatic charge in air. For many years this unit was the roentgen (R), which was defined as the amount of radiation that produced one “electrostatic unit of charge” (about  $2 \times 10^9$  ion pairs) in  $1 \text{ cm}^3$  of air at standard temperature and pressure. This does not mean much, even to physicists, because these are not natural or logical units. Since 2006, the use of R has been discouraged and the preferred unit is now coulombs per kg air (C/kg air), which has the advantages of being internally consistent in the SI system and meaningful to physicists; it is also scaled appropriately to human experience. One C/kg air is approximately equivalent to 3876 R, and one R is approximately equal to  $2.58 \times 10^{-4}$  C/kg air. A typical annual exposure for an individual who is not occupationally exposed or living in an area of unusually high background radiation is  $200 \text{ mR} = 7.8 \text{ C/kg air}$ .

However, the amount of radiation that reaches the object or person is not as important to an OEM physician faced with a radiation-related issue as is the amount absorbed and especially the effective dose.

Measurements of dose reflect the total quantity of radiation absorbed by the person or object. The “rad” was a standard measurement of absorbed dose indicating the deposition of energy in the material

absorbing the radiation (100 erg/g); roentgen (R) measurements were treated as equivalent to rads in the regulations of the U.S. Nuclear Regulatory Commission, although that is not strictly correct. The “gray” is the SI redefinition of the same thing using preferable SI units (1 Gy = 1 joule/kg), so that 1 Gy = 100 rad.

The amount of radiation absorbed by an object or person means little by itself, however. In order to quantify the effect on materials and on living tissue the absorbed dose needed to be modified by some factor representing the physical characteristics, or “quality” of the interaction, which is a function of the LET (described above). The “rem” was a modification of the rad measurement, taking into consideration the “relative biological effect” (RBE) induced by the type of radiation (in other words, the biological damage that could be caused by the LET phenomenon). Rem was a measurement of the absorbed dose adjusted by the effect of the type of radiation in tissue and was therefore very important in dosimetry, or estimating health risks. Just as the gray was a recalibration of the old rad measure, the SI replacement unit for dose equivalence, the sievert (Sv), is a recalibration for the rem: 1 Sv = 100 rem.

“Working level” (WL) is an administrative unit used in uranium and other underground hard-rock mining in which radon and other radionuclides are present. It is a quantity of radiation from exposure to radionuclides in any combination that results in a certain total quantity of alpha particle emission energy (130,000 MeV). Standards and guidelines for work underground were based on the cumulative length of time a miner was exposed at this level, usually expressed in “working level months” (WLM). One WLM equals exposure to 1 WL for one month (720 hours), with average environmental exposure contributing about 4 WLM of background exposure. WLM has been useful in defining the maximum time a miner was allowed to spend underground in a particular mine, and was adopted out of necessity for epidemiological studies of radiation-exposed miners but is not very useful for any other purpose.

The actual measurements in occupational health surveys are made using different technologies, among them ionization of air, ionization and production of current in a detector tube (such as the familiar

Geiger counter), scintillation counting, electron entrapment and release from thermal heating, and interactions with photographic film emulsions. Special techniques are used in high-energy physics and the nuclear industry, to cope with much higher energy fluxes.

Measurement of the personal exposure of workers to radiation is universal in healthcare and in industries in which a high level of radiation hazard exists, but is uncommon in mining operations and situations where only low-level exposure is expected. Film badges, thermo luminescent dosimeters, and pocket ionizing chamber dosimeters are most often used for assessing individual worker exposure.

Certain radionuclides present particular shielding problems.  $^{32}\text{P}$ , for example, is a beta emitter that produces x-rays passing through lead and must therefore be shielded with thick plastic.

## Health Effects

External  $\gamma$  radiation, such as the beam of an x-ray machine, is familiar to most physicians. The energy level of the beam is the most important determinant of risk. High levels of exposure may be associated with erythema, scarring, and skin atrophy, unfortunately familiar to physicians as side effects of radiation therapy. External radiation damages skin as a gradient phenomenon and induces cancer as a stochastic phenomenon (see Chapter 2), so that the local injury at the site of radiation is worse with increasing exposure and the risk of cancer resulting from exposure is increased.

External  $\alpha$ ,  $\beta$ , and neutron exposures are not an occupational issue outside the nuclear and nuclear weapons industry. External  $\alpha$  radiation does not penetrate skin and therefore has little effect. External  $\beta$  emissions penetrate variably depending on the velocity with which they are emitted; most do not penetrate enough to cause injury but high energy  $\beta$  can damage skin.  $\beta$  can be shielded with thin aluminum.

The effects of internal radiation depend on the deposition, absorption, distribution, and retention of the radionuclide in question.  $^{131}\text{I}$  is heavily concentrated in the thyroid and is strongly associated with an increased risk of thyroid cancer; immediate administration of stable

iodine to block uptake of  $^{131}\text{I}$  is a mainstay of treatment when exposure to this isotope is suspected.  $\alpha$  radiation is most often associated with mining in the context of occupational and environmental medicine because inhaled particles may contain  $\alpha$ -emitting radon daughters (decay products of radon, several of which emit  $\alpha$  radiation themselves). Emissions from radon daughters are known to increase risk of lung cancer among underground miners, especially uranium and phosphate miners.  $\alpha$  emitters, such as  $^{239}\text{Pu}$  and Ra isotopes, Rn and its “daughters,” and  $^{210}\text{Po}$ , induce heavy exposure in a small circumscribed area of tissue, especially the immediate area surrounding a particle that is deposited on the bronchial epithelium, and are associated with an elevated risk of lung cancer following inhalation.  $^{238}\text{Pu}$  and other Pu isotopes, once absorbed, are also concentrated in bone, where they may be associated with the risk of osteogenic tumors. The  $\alpha$ -emitting isotopes most likely to be of significance in occupational medicine outside the nuclear industry are radon and its daughters.  $^{210}\text{Po}$ , on the other hand, is found in cigarette smoke.  $^{32}\text{P}$ , a  $\beta$  emitter used in some nuclear medicine applications, is distributed widely in the body, exchanging freely with intracellular phosphorus.

The health effects of ionizing radiation can be classified as non-stochastic (predictable for an individual, as in the case of cataracts and acute radiation sickness) and stochastic (probabilistic, predictable for the population but not the individual, such as cancer).

Cancer risk dominates the issue of low-dose exposure to radiation. A less-studied outcome is cardiovascular disease, for which there is evidence of a weaker association.

Although far less important in practice than their radiation emission properties, radionuclides may also have intrinsic toxicity, just as do their nonradioactive counterpart isotopes. Uranium, for example, is nephrotoxic and this may be of significance in unusual exposure settings.

### ***Cancer Risk***

Exposure to external radiation is a known risk factor for certain types of cancer. Leukemia as a class, except for chronic lymphocytic

leukemia, is well known to be associated with exposure to ionizing radiation. Stomach, lung, thyroid, and breast cancers are also elevated following exposure but not as consistently or as markedly as leukemia. Much of the available information on cancer risk comes from follow-up studies of patients treated for cancer, who may be predisposed to cancer, and survivors of the atomic bomb explosions in Japan at the end of World War II. There are fewer studies that have characterized cancer risk from occupational exposure.

The International Agency for Research on Cancer (IARC, see Chapter 2) identified several radiation-induced cancers and has quantified their risk following low-level radiation. In an ongoing international collaborative study that now includes 600,000 radiation-exposed workers in the nuclear industry in 17 countries, IARC investigators found a relative risk for all cancers, excluding leukemia, of 0.97 per Sv. It was concluded that at the average dose observed in the study (19 mSv) this increased exposure would correspond to a 2 percent increase in overall risk of dying from cancer, excluding leukemia. For leukemia, other than the chronic lymphocytic type, they found a much higher increased risk: 1.93 per Sv, although there was much uncertainty. The effect was unlikely to be confounded by smoking. Atomic bomb survivors showed an increased risk of 0.32 per Sv for cancers excluding leukemia, and 1.54 to 3.15 per Sv for leukemia, excluding chronic lymphocytic.

The risk from low-level exposure to external radiation has been highly controversial. There has been a substantial reevaluation of estimates of the risk of low-level ionizing radiation as more reliable data have become available, mostly in the form of more accurate estimates of dosage among Japanese survivors of either atomic bomb explosion, increased numbers and better documentation of patients undergoing medical irradiation, and animal experimentation. The findings of this reevaluation are summarized in a series of documents called the BEIR Reports (for “biological effects of ionizing radiation,” fifth revision), prepared by the U.S. National Research Council. In recent years these reports have taken a more conservative position than in the past, associating health effects with generally lower levels

of radiation than in the past. BEIR V suggested that the mutation rate in humans is likely to be doubled at about 1 Gy. Leukemia emerges as the cancer most sensitive to radiation level; other cancers appear to be elevated only at higher dosage levels and follow an approximately linear dose-response relationship. Mental retardation is also recognized as a significant effect following exposure in utero. As a result of these findings, occupational exposure standards for radiation are undergoing reevaluation and are likely to be reduced in the future.

The risk of internal exposure, from ingested or inhaled radionuclides, was well studied in the early years of the nuclear industry in animal studies. Plutonium is an  $\alpha$  emitter that, once inhaled or ingested and distributed, is incorporated into bone. Once there, it irradiates the osteoclastic cells in close proximity, inducing osteogenic sarcomas. In the 1920s, radium was used to paint luminescent watch dials. Workers who did this, mostly young women, tipped the brushes in their mouth, which resulted in a heavy exposure to this  $\alpha$  emitter and a high risk of osteogenic sarcoma. (A very similar but toxicity-based outbreak occurred in the nineteenth century involving elemental phosphorus.)

Natural thorium in rock decays to radium and then to radon, which, being a gas, is free to escape through cracks or dissolve in water. Radon adsorbs readily to surfaces, including glass, and to the surface of particles suspended in air. It accumulates in poorly ventilated spaces where there is free communication with rock or underground water, such as unfinished basements, and therefore presents an indoor environmental hazard. (This is discussed further in Chapter 11.) Airborne radon and its “daughters,” the products of its decay, are  $\alpha$  emitters. Inhalation of a particle with adsorbed radon carries the radionuclide to the deep lung, where it deposits and delivers local irradiation to surrounding cells in a small volume. This may induce bronchogenic carcinoma, which has been shown to have an interactive risk with cigarette smoking in uranium miners. Radon and its daughters are thought by the U.S. EPA to be the second leading cause of lung cancer, after cigarette smoking (although asbestos may be of similar magnitude), and together with

passive cigarette smoke exposure it is the leading cause among non-smokers.

### ***Acute Radiation Syndrome***

Radiation sickness is the colloquial name for “acute radiation syndrome,” which results from intense whole body radiation. It may occur after external beam irradiation or internal absorption of a radionuclide. Recovery should be expected after an exposure of <6 Gy but is usually fatal at 10 Gy without medical treatment, or at 12 Gy despite medical treatment, and is rapidly fatal above 30 Gy.

The earliest symptoms occur within minutes and include nausea, vomiting, and diarrhea, and lymphopenia. This lasts for hours (except in overwhelming radiation exposure), and the patient seemingly then returns to his or her previous baseline. Acute radiation syndrome is characterized by this latent phase, which may last several days.

As the syndrome progresses, widespread cellular injury results in progressive multisystem disease, beginning with those tissues in which cell turnover is most rapid. Cell replacement is inhibited by the radiation-induced genetic damage, and failure of cell proliferation results in mucosal ulceration, skin lesions, leukopenia, thrombocytopenia, and subsequently in alopecia and anemia. Clinically, the most severe manifestations are usually gastrointestinal bleeding, infection in the presence of immune compromise, cerebral edema, and profound pancytopenia. Survivors are at risk for dementia, cancer (particularly acute leukemia), cataracts, and infertility.

Immediate decontamination of persons exposed in radiation incidents involving the internal absorption or implantation of radionuclides can sometimes be life saving and can significantly reduce disability. This generally involves immediate and copious washing of skin avoiding abrasion or excoriation, wound debridement, excision of embedded particles from tissue, lung lavage, gastric lavage, and nasal cleansing, as indicated by the nature of the incident. Facilities in

which the potential exists for such exposures must be equipped for these procedures and must have contingency plans for disaster management.

Obviously, decontamination measures are not needed if the source of exposure is external, such as inadvertent exposure to a beam. The priority in cases of external irradiation should be supportive care. Limited or segmental external radiation may produce skin erythema at the site of exposure, tissue necrosis, and variable risk of cancer with high exposure. Specific effects of local irradiation may include thyroid cancer with irradiation of the neck (relatively rare), cerebral edema and dementia with heavy cranial irradiation, and birth defects, particularly neurological, with irradiation in utero at relatively low levels.

### ***Regulation***

Regulation of exposure to radiation is somewhat inconsistent.

The OSHA standard for occupational exposure to ionizing radiation uses units of rem per calendar quarter (three months) out of administrative convenience. OSHA regulations are concerned almost exclusively with external radiation because intake of radionuclides would be very rare outside the nuclear or nuclear weapons industries. It sets allowable exposure at 1.25 rem/quarter (0.015 Sv/quarter) to the whole body or gonads, 7.5 rem/quarter (0.075 Sv/quarter) to the skin of the whole body, and 18.75 rem/quarter (0.1875 Sv/quarter) for the hands and forearms, which of course are away from radiosensitive internal organs. To reflect reduced susceptibility with age, the accumulated dose to the whole body over the entire time at work (not just one quarter) cannot exceed  $5(N - 18)$  rem ( $5[N - 0.18]$  Sv), where  $N$  is the worker's age at his or her last birthday, and may not exceed 3 rem in any calendar quarter. Workers under eighteen years of age cannot be exposed to more than 10 percent of the accumulated limit (which would equal  $0.5(N - 18)$  rem in any calendar quarter).

The Department of Energy regulates the nuclear industry and has its own set of standards for occupational exposure to ionizing radiation. Because the nuclear and nuclear weapons industries may involve

incidents of release, DoE is concerned with both external radiation sources, such as beams and radioactive materials, and internal doses from ingested or inhaled radionuclides. DoE also uses the term "committed equivalent dose" to refer to the sum of the estimated doses received by an internal organ of radionuclide which may have entered the body. This is calculated using the sum of estimates of how much radiation reached each organ, weighted by a coefficient representing the relative sensitivity of that organ to radiation, over a period of fifty years. Lung, bone marrow, and the gastrointestinal tract are given the highest weights. DoE sets an annual exposure limit of 0.05 Sv, distributed by target tissue or organ and calibrated as follows:

- The allowable exposure to the lens of the eye is 0.15 Sv (assumed to penetrate 0.3 cm or less).
- The allowable exposure to the skin or to any extremity in a shallow dose is 0.5 Sv (assumed to penetrate 0.007 cm or less); skin is subject to further qualifications depending on the size of the area irradiated.
- The sum of the deep-dose equivalent for external exposures (taking the penetration factors above into account) and the committed-dose equivalent to any one organ or tissue (other than the lens of the eye) is 0.5 Sv (which, in effect, sets a maximum allowable exposure for any one organ).
- The allowable exposure to the fetus in a "declared pregnant worker" in the nuclear industry is 0.005 Sv, with numerous further provisions for preventing further accumulation of exposure if this is exceeded.
- The allowable exposure to visitors to facilities, who are not employed in the industry, is 0.001 Sv.

These basic exposure standards are supplemented by extensive requirements and prescribed work practices. OEM physicians who work in the industry will as a matter of course become intimately familiar with them.

Exposure from background, therapeutic (radiation therapy), research-related (voluntary), and diagnostic medical radiation are not included in these standards.

The Nuclear Regulatory Commission is the responsible federal agency in the United States for all licensed users of nuclear materials (such as radionuclides in nuclear medicine) or operators of nuclear reactors. These regulations govern many aspects of work in facilities handling radionuclides and in nuclear plants, including mandatory fitness-for-duty programs and mandatory drug screening.

The U.S. Environmental Protection Agency is involved in planning for radiological emergencies. There are two basic regulations: one for emergency contingency planning in the event of a significant release of radioactivity (in collaboration with the Federal Emergency Management Agency), and the other on emergency planning for release of radioactivity from a commercial nuclear plant. In 2008, under pressure to establish standards for waste nuclear reactor fuel disposal at the Yucca Mountain repository, the EPA promulgated standards for nuclear waste disposal, which would allow 15 mrem (0.015 mSv) per year for the first 10,000 years after disposal, followed by 100 mrem (1 mSv) per year for the next million. How this unfunded mandate will be enforced by future civilizations is unclear.

## **NONIONIZING ELECTROMAGNETIC RADIATION**

Nonionizing radiation consists of electromagnetic radiation at energies too low to ionize atoms in matter; unlike ionizing radiation, no particles are involved other than photons. The electromagnetic spectrum below about 100 nm can be divided into several spectral ranges, described by wavelength or frequency: ultraviolet light (100–400 nm), visible light (400–700 nm), infrared (700 nm–300  $\mu$ m), microwave (0.1 cm–10 m), and “long” radiofrequency waves (>10 m). They are discussed here in order of their place on the electromagnetic spectrum.

## **Ultraviolet Radiation**

Ultraviolet radiation (UV) is a common and serious occupational hazard ubiquitous in outdoor work, and common in certain occupations such as welding, in which radiation is produced by open electrical arcs. There are three components to the UV spectrum: UVA (400–320 nm), UVB (280–320 nm), and UVC (280–100 nm).

UVA is often called “long-wave,” “near UV,” or “black light.” It is relatively low in energy and weak in eliciting sunburn or other local effects. UVB is much more efficient in producing skin erythema, particularly at a wavelength of 297 nm. It is also capable of penetrating light skin to a depth of a few millimeters, reaching cells in capillary blood and the dermis. UVC is highly destructive to both bacterial and mammalian cells, and is useful for its germicidal properties, but it is readily absorbed by air and thus cannot propagate far in the atmosphere. UVC reaching the Earth in sunlight is absorbed by the stratospheric ozone layer so that what reaches the surface of the Earth is UVA and UVB radiation. Conventional sunscreens, such as para-aminobenzoic acid, tend to absorb and protect against UVB most efficiently. (Zinc oxide and new sunblocking products containing nanoparticles of titanium oxide block UV efficiently at all wavelengths.)

Exposures to UV of concern in occupational medicine include all occupations in which there is prolonged activity out of doors, regardless of temperature or climate. Severe UV exposures may occur even in cold northern conditions, aggravated by reflection off snow or water. Overcast conditions can be misleading, as UV is essentially unaffected by cloud cover although visible light may be filtered out. Intense UV exposure may also occur in welding and metal cutting operations, particularly gas metal arc welding, plasma arc welding or cutting, carbon arc welding, and welding on aluminum. Ultraviolet light sources also have some uses in sterilization and the production of special effects using fluorescent paints.

Exposure to UVB (280–315 nm) is associated with sunburn, a response that reflects cell injury in the basal cell layer and may result in blistering and even systemic symptoms. The degree of erythema

produced depends directly on exposure. How slowly or rapidly erythema develops depends on attenuation by skin pigmentation, the thickness of the stratum corneum, and the presence or absence of photosensitizing chemicals or photo-absorbing sun blocking agents. On recovery, the cells of the basal layer secrete melanin to increase protective pigmentation, which is the basis for tanning.

Skin cancers (squamous cell, basal cell, or melanoma) are associated with exposure to UVB. This strong association is increased in the presence of photosensitizing chemicals, as used in the roofing industry, where workers are exposed to polycyclic organic hydrocarbons. Squamous and basal cell carcinomas are relatively slow to grow and late to metastasize; they are a cosmetic nuisance but are easily excised in most cases. Melanoma, on the other hand, may be a highly malignant lesion and often metastasizes early. Melanoma appears to have a more complex relationship to UV exposure, as this often highly aggressive skin cancer appears to be related to sun exposure in incidence, but its distribution on the surface of the body does not correlate closely with that expected from sun exposure with the usual clothing. There is increasing evidence to suggest that melanoma is associated with infrequent but severe sunburn, especially early in life. There is also evidence suggesting a more complex process, involving inhibition by vitamin D synthesis. Because vitamin D is synthesized in skin in response to UV, the relationship, if there is one, is complicated. UV does not penetrate skin and so exposure does not induce cancer in deeper structures.

Other skin problems that have been reported to be associated with UV exposure, particularly among welders, include discoid lupus and polymorphous light eruption, a persistent diffuse erythematous reaction often seen on the face. Exposure to UV may also aggravate existing photosensitization due to porphyria cutanea tarda (which may be associated with alcohol abuse), systemic lupus, medication, or other causes unrelated to occupation.

Eye problems associated with exposure to UV include cataract, keratitis and conjunctivitis, and retinal damage. Cataracts tend to occur after intense exposure over many years. This effect has been particularly

well documented among Chesapeake Bay watermen. Keratitis is a reaction that is intensely painful because of the numerous sensory endings on the corneal epithelium; it is incapacitating while it is present. It usually begins soon or a few hours after exposure and is associated with a generalized conjunctivitis. This is often called "welder's flash" or "flash-burn." It may persist for several days, although it resolves without permanent damage. Recovery from welder's flash does not provide future protection, as melanotic pigment cannot be produced by the cornea. Retinal injury is a potential problem for a narrow band of UV, 310–340 nm, to which the retina appears unusually sensitive. Although the lens focuses UV on the retina, most of it is absorbed by the lens and cornea, where the major damage is produced.

Immunological changes have also been induced by exposing animals to UV. It appears that cellular immune responses can be blunted. Whether this is a practical problem in occupational exposure to UV remains to be demonstrated.

Protection against the effects of UV lies in avoiding exposure. Sunscreens and blocking agents, although popular for sunbathing, are not adequate for the intense and sometimes prolonged exposures that typically occur in occupational situations, and there has been concern over the safety of some preparations. Control measures in the workplace include using nonreflective paints (containing titanium dioxide or zinc oxide) on surfaces that are placed in such a way as to reflect UV toward workers, hanging UV-impermeable plastic curtains between workstations, and segregating activities such as welding to enclosed booths. Personal protection can be achieved by protective work clothes and by viewing the welding arc through helmet visors or hand shields with filter lenses. These lenses are rated by number, and tables are available from the American Welding Society giving the recommended minimum and suggested lens shades for different welding operations.

### **Light in the Visible Spectrum**

Light in the visible spectrum spans a very narrow range, 380 to 750 nm. These energy levels are too weak to ionize atoms but carry

sufficient energy to activate the rods and cones of the eye and to power photosynthesis in plants. The eye and brain perceive increasing wavelength within the spectrum as color, in the rainbow progression of violet, blue, green, yellow, orange, and red.

There are three issues in occupational medicine that involve light in the visible part of the spectrum: adequacy of illumination, the psychological effects of variable lighting, and the health hazards of coherent light from laser beams.

Illumination is a fundamental requirement of almost all work standards, and the adequacy of illumination is a limiting factor for accuracy and safety. Inadequate lighting is associated with headaches and fatigue, as considerable effort must be exerted to keep details of work in focus, but there is no long-term damage to the eye from squinting or eyestrain. As the worker ages, accommodation becomes more difficult, and increased light levels are required to see and resolve small or distant objects. These changes make illumination particularly important for older workers. It is also more difficult for older people to adapt to changes in illumination, as when entering a dark room from the sunny outdoors or seeing through glare. The productivity and comfort of younger workers also suffers when illumination is inadequate. Standards for lighting are specified for a limited number of situations by OSHA and for general purposes by the American National Standards Institute. Illumination levels should be adequate to meet the needs of work, regardless of the standard. The smallest detail of work should be easily resolved visually, there should be sharp contrast between the background and the objects worked on, and sources of glare should be identified and removed.

Cycles of light and dark are environmental cues that reset biological rhythms of sleep, alertness, and mood. Disturbances in light/dark cycling interfere with these endocrinologically and neurologically mediated rhythms and may result in sleep disturbances (usually insomnia), reduced alertness, and depression. This is a particular problem in the form of jet lag, when combined with other features of long-distance travel. It occurs in shift work as well as in work in prolonged relative darkness, as in underground mining and work in

northern latitudes. When associated with seasonal changes in lighting, this syndrome has been given the name "seasonal affective disorder" (SAD). It is treatable with artificial illumination similar to the spectral composition of sunlight. Other situations of light/dark cycle privation can be managed by installing windows and skylights, or by using strong artificial lighting simulating sunlight. Commercial headsets containing light sources are available for the same purpose. These headsets are mounted on the forehead but are positioned so they do not shine directly into the eyes. (See Chapter 13.)

Coherent light, in the form of laser beams, carries much more energy over a much longer distance than diffuse visible light; it excites the atoms and molecules it encounters but does not ionize them. This property makes coherent light useful in applications such as surgery, materials cutting, measuring instruments, and signal transmission. When focused by the lens on the retina, a laser beam can cause severe retinal injury and permanent destruction at the focal point on the retina. Local injury to the external structures of the eye and skin is possible with lasers carrying high amounts of energy. Safeguards should be present on all lasers that are not completely contained so that the lasers cannot be inadvertently tipped toward observers and heads, hands, or other body parts cannot move into the beam. Enclosures for the beam should be constructed of nonreflective, fireproof materials. Reflective surfaces must be removed from the area or covered, as the reflected beam usually carries almost as much energy as the original beam. Training in the safe use of lasers is mandatory in any situation where a worker could be exposed to the beam. Lasers are also produced that emit in the ultraviolet and infrared spectrum; the effects and hazards vary somewhat with the wavelength, and are similar to the effects of non-coherent electromagnetic radiation in the same frequencies.

### Infrared

Many industrial processes use furnaces, thermal sealing devices, stoves, heat lamps for baking or drying, or any of numerous other

heat sources that emit infrared radiation. In occupational medicine, exposure to infrared radiation may be a problem in any occupation or industry in which hot processes play a role, including foundries and steel mills, glassblowing works, heating and dehydrating operations, and welding.

Infrared radiation is produced from all objects to a degree depending on the temperature. Radiation heat loss is a major cooling mechanism for the human body in which infrared radiation is absorbed by bodies that are cooler than the source. How much heat an object will take on depends on its composition and the wavelength of infrared to which it is exposed. The change in temperature that results when an object is heated is described by the “specific heat” of the substance and the mass through which the heat diffuses. Formally,

$$\Delta T = Q/cm$$

where  $Q$  is the heat transferred to the object,  $c$  is the specific heat of the material of which it is composed, and  $m$  is the mass of the object. Metals have a very low specific heat and warm very quickly. Water has one of the highest specific heats of any substance and absorbs proportionately much more heat as the temperature rises. Changes in phase also involve heat exchange. As water turns from ice to liquid or from liquid to steam, it absorbs even more heat as it changes phase without changing temperature. Therefore, perhaps counter-intuitively, the greatest damage to the body occurs when it is exposed to infrared at wavelengths that transfer heat but are not efficiently absorbed by water (750 to 2000 nm), but not at higher wavelengths, which carry more energy but are more efficiently absorbed by body water (2000–3000 nm). The slower rise in temperature at that wavelength due to absorption by intra- and extracellular water attenuates the thermal effects of infrared radiation. (This is in contrast to microwave radiation, in which it is precisely the thermal effects of water heating that leads to injury at a particular wavelength.)

Absorption of the infrared radiation results in thermal effects on the body, and the most vulnerable target tissues are superficial structures, such as the skin, cornea, iris, and lens. Infrared does not penetrate well; therefore, deep structures are usually not affected by infrared itself except in the most extreme situations.

Exposure to infrared radiation may result in erythema and eventual blistering and burning of the skin if acute pain does not result in withdrawal from exposure. Exposure to the eye over prolonged periods may result in cataracts (often referred to as “glassblower’s eye”) and potentially severe retinal injury. In situations where there is heating of the total body, heat stress may be a significant hazard. (Heat stress is discussed elsewhere in this chapter.) There is also a risk of thermal burn in situations of exposure to infrared radiation. Microwave radiation, discussed elsewhere in this chapter, also presents heat-related hazards, but these effects are deep and do not affect the skin.

## Microwave

The term “microwave” refers to radiofrequency radiation on the spectrum below infrared but above conventional radio and television transmission. The part of the spectrum of greatest significance in OEM includes the wavelength of 1 cm, which is the most efficient for heating water and therefore has the greatest thermal effect in biological tissues. Microwave transmissions are used for radar, communications beacons, heat-sealing, cooking, diathermy, and plywood bonding, among other applications. The microwave beam can be highly focused and reflected, and directly excites molecular movement, causing heat. Microwave irradiation from a strong source induces deep heating, particularly in water-rich materials such as tissue or food.

The possible effects of microwaves on human health are controversial. There is a single documented case of fatality in a heavily exposed worker who stood in front of a radar transmitter; moderately heavy exposure has been common in the past, however, and no

strong pattern of health consequences is evident. The best-documented effects include cataract formation and the demonstration in animals of male reproductive effects as a consequence of the thermal heating. Reports of mental confusion, auditory perception (a hissing or clicking sound), lightheadedness, fainting, and, in animals, fetal death and anomalies have been reported. Immunological and hematological abnormalities have been suggested. Research in this area was a particular specialty of Soviet occupational health investigators, who developed a considerable literature describing behavioral effects that Western scientists have found difficult to replicate. This discrepancy explains in part why occupational exposure standards for microwave may differ by two or three orders of magnitude among various countries. Radiofrequency radiation in this range, as in the case of microwave ovens, may also disturb the function of cardiac pacemakers.

In general, the effects appear to be related to exposure level and can be controlled by shielding and preventing exposure to the beam. The effects appear to be related to frequency and pulse as well as to intensity.

Recently there has been speculation that the use of cell phones, which operate at 900 MHz at rather low power levels, may be a risk factor for brain tumors, based on purported associations between the use of handheld devices on one side and brain cancer on that side. This concern has stimulated much research, which on balance has not documented an association. The evidence for an association is extremely weak and is not widely accepted. The signal strength and magnetic fields associated with cell phones are thought to be too weak to affect biological systems.

### **Radio Frequencies below Microwave**

Radio frequencies (RF), characterized by long wavelengths, are usually described in terms of frequency for convenience. Radio frequencies overlap and fall below the microwave range (centimeters) on the spectrum, and include “very high frequency” (VHF, 30–300 MHz,

wavelengths on the order of a meter), “high frequency” (HF, 3–30 MHz, 10 meters), “medium frequency” (MF, 0.3–3 MHz, 100 meters), “low frequency” (LF, 30–300 kHz, 1000 meters), “very low frequency” (VLF, 3–30 kHz, 10,000 meters), “ultra low frequency” (ULF, 0.3–3 kHz, 100,000 meters), and “extremely low frequency” (ELF, 3–300 Hz, 1 million meters). VHF is familiar from television transmission. HF includes frequencies used for citizens band radio, FM radio, amateur and international radio, and emergency channels. MF includes bands for AM radio transmission. Lower frequencies are used for power transmission, heat treating, and very-long-distance broadcasting. ELF frequencies penetrate very deeply into the Earth and are used for submarine communications and other defense applications.

At the field strengths encountered in normal industrial applications, RF does not appear to pose a substantial hazard. When effects have been suggested, they have been similar to those mentioned for microwave radiation. ELF has been particularly controversial as a result of its defense applications, and claims of increased cancer risk have been advanced but are unsubstantiated at this time. As is the case for microwave, shielding and avoiding exposure to the concentrated beam seem to provide protection.

## NOISE

Noise is sound energy—the propagation of vibration, or waves, in air. This form of energy, like electromagnetic radiation, can be characterized by frequency ( $\nu$ , expressed in vibrations per second or “hertz,” abbreviated Hz), wavelength ( $\lambda$ , which is the inverse of frequency), amplitude of the wave, perceived loudness (which is a function of amplitude and perception), the energy content of the wave motion, and periodicity. Noise may be “continuous” over a period of time or may be “impulse” noise, which is dominated by a single high-energy pulse, such as a blast, the two types having very different health effects.

The loudness of a sound reflects the magnitude of pressure of the sound wave propagated in air. The loudness of sound is measured in “bel” (B), named after Alexander Graham Bell. The bel scale is

exponential: 4 B is 10 times louder than 3 B and 100 times louder than 2 B. A bel is an inconvenient unit to use for most purposes, so it is customarily divided into tenths, or “decibels”; a 40 dB noise is equal to 4 B and is 10 times louder than a 30 dB (3 B) noise. Because the scale is exponential but divided into tenths, the logarithm of doubling is about 3 dB; that is, a 43 dB noise is about twice as loud as a 40 dB noise (Table 9.1).

Human beings with normal hearing hear differently at different frequencies. A pure tone at 1000 Hz sounds louder than a pure tone carrying the same energy at 250 Hz. In order to accurately adjust for how the hearing mechanism and the brain respond to different frequencies, physical measurements of sound using the decibel measurement are modified by electronically varying the sensitivity of the measurement device to the frequency components of the noise, using a formula called the “A-scale.” Devices that measure continuous sound are called “sound level meters” (SLM) and are visually calibrated for occupational monitoring purposes to read automatically in “dBA,” decibels weighted by the A-scale. (There are other scales with other applications.) Devices that measure sudden, direct noises are

**Table 9.1.** The Decibel Scale

Relative Loudness	Bels	Decibels
$10^{11}$	11	110
$10^{10}$	10	100
$10^9$	9	90
$10^8$	8	80
$10^7$	7	70
$10^6$	6	60
$10^5$	5	50
$10^4$	4	40
$10^3$	3	30
$10^2$	2	20
$10^1$	1	10
$10^0$	0	0

called "impulse meters." Those that measure the total exposure to noise over a given time are called "noise dosimeters."

Table 9.2 gives examples of dBA readings that one might encounter in common situations. In practice, SLM readings are usually made in the workplace by safety officers, industrial hygienists, audiologists, or acoustical engineers, and in the community by noise control officers representing local government. Exposure to noise in the workplace occurs in jobs in many industries and depends on the process being performed, the acoustics of the workplace, the controls on the machines or noise sources, the distance between the source and the worker, and what personal ear protection the worker may be using. Exposure in the community is usually less intense but may be increased for residents of communities near major airports, highways, or industrial plants, or for individuals who indulge in motorcycle riding, shooting firearms, or listening to music at high volume.

Because the bel scale is exponential, and the decibel scale is an arithmetic division of an exponential scale into tenths, the decibel scale has some peculiar properties when used in practice. For example, a sound that registers 40 decibels is not twice as loud as a sound that registers 20 decibels. It is 100 times as loud, because the difference (20 decibels = 2 bels) represents two orders of magnitude,  $10^2$ . It is easy for the OEM physician to forget this and fail to appreciate the profound difference in noise intensity that a few decibels may represent. If two equal sources of noise are placed together, the sound

**Table 9.2.** Noise Levels in Typical Exposure Situations

Noise Source	dBA
Chainsaw	110
Jet flying overhead (1000 feet)	100
Motorcycle (25 feet away)	90
Diesel truck (50 feet away)	85
Garbage disposal	80
Television set	70
Office with air conditioning	60

energy will double, but this will equal about 3 dB (0.3 bel, a factor of  $10^{0.3}$ , or roughly 2). Confusingly, some occupational health regulations advocate a doubling factor of 5 dB, which is simply incorrect. The sound at the action level (85 dB) is actually less than the permitted exposure level, or PEL (90 dB), by a factor equal to the square root of 10 ( $10^{0.5}$ , or about 3.2)—a substantial difference from doubling but one that serves to protect the worker.

Noise-induced hearing loss can be prevented by avoiding exposure to excessive noise. This requires a program of noise control as well as monitoring workers for early detection of hearing loss (Table 9.3). Noise control is a highly technical specialization that may involve acoustical engineering, plant design, engineering controls, and containment or isolation of noise sources. However, most problems involving excessive noise can be handled effectively and inexpensively using basic principles.

The OSHA Noise Standard in the United States requires employers to measure the noise level in the workplace if the noise is loud enough to interfere with easy speech communication or to cause discomfort. If the noise level exceeds the eight-hour time-weighted average (TWA) of 85 dBA (the “action level”), then the employer must institute a hearing conservation program. The PEL is 90 dBA (8-hour TWA).

Hearing conservation programs should include regular monitoring of the workplace, baseline and annual audiograms for all exposed workers, in-service and pre-service worker education regarding hearing conservation, systematic record-keeping and worker notification when problems are detected, and the provision of hearing protection to all exposed workers. Many programs in industry also

**Table 9.3.** Components of Hearing Conservation Programs

Noise monitoring in the workplace	Acoustical noise abatement
Training of new employees	Updated training
Annual audiometric evaluations	Employee notification and referral
Personal hearing protection	Systematic record-keeping

include referral of affected employees to specialists, administrative controls to limit the duration of assignments in noisy areas, and noise control measures. Noise-induced hearing loss is discussed in detail in Chapter 17.

The weakest part of most industrial hearing conservation programs is personal protection. To be effective, a variety of effective hearing protection devices must be made easily accessible to all exposed workers, who also must be trained in their proper use and motivated to wear them. Hearing protection devices vary considerably in their effectiveness and are rated by their ability to attenuate sound. A device that is overly impermeable to sound for a given job could be a safety hazard if it interferes with communication. In practice, this seldom occurs and workers usually have no problems understanding speech once they become accustomed to the device. Individuals vary considerably in their ability to tolerate various models of earmuffs and plugs, and a selection must be available for workers to choose what they prefer. Workers must be instructed in their correct use: nondisposable earplugs must be kept scrupulously clean, deformable-type disposable plugs should never be inserted with dirty fingers, and earmuffs should be seated properly on the head. If the devices are uncomfortable, inconvenient, inaccessible, or ridiculous-looking, some workers will simply not use them or will remove them whenever they are unsupervised. Rolled-cotton earplugs are often used, but are almost completely worthless because of their poor sound-absorbing characteristics. Even worse, they may provide a false sense of protection.

## PRESSURE AND ALTITUDE

Barometric pressures above or below 1 atmosphere are part of the conditions of work in special environments. Pressure is measured in terms of force per area, using various units, and in physics is often called "force density." In occupational and environmental medicine, one is generally most interested in pressures near 1 atmosphere (1 atm). That is the pressure at the surface of the Earth, to which

human beings are adapted. If the pressure is much below this level, then the human body experiences decompression effects; if it is much above this level, then the human body becomes compressed or even crushed.

In scientific and engineering usage, the standard unit is the pascal (Pa), but this is inconvenient for most other purposes. One atmosphere is equal to 101.325 kPa, 1 kg/cm<sup>2</sup>, or 760 torr (previously “mm of mercury” and of historical interest because Torricelli first measured atmospheric pressure using a mercury barometer), which in turn is almost but not quite equal to 1 bar. The “bar” was derived for meteorological use and is equal to 10<sup>6</sup> dyn/cm<sup>2</sup>, and is another unit more convenient in scientific applications than in medicine. In industry, where the conventional English system is still often used, 1 atm equals 14.7 pounds per square inch.

Absolute pressure is usually less critical than differences in pressure and the changes experienced by the worker. Direct adverse effects of pressure changes are called barotraumas and are of particular interest in aerospace and undersea medicine. There are also a number of problems that result indirectly from pressure changes, mostly from gases dissolving into and then being released out of body fluids. The literature is highly specialized and technical and is firmly grounded in physiological principles. Generally in occupational medicine, problems associated with pressure changes are relatively uncommon except in the context of commercial air travel or shallow diving. In military occupational or operational (combat) medicine, such problems are common and critically important.

## **Compression**

Descent into a body of water increases pressure on the body by approximately 1 atmosphere per 10 m vertical distance. Problems may arise with either compression (going down) or decompression (coming back up).

The internal pressure in the body equalizes or adapts to increased external pressure relatively rapidly as gas dissolves into tissue, and air

in communicating spaces in the body is compressed. Health problems associated with compression are relatively uncommon at depths where industrial activity is likely to take place. However, decompression effects are common and can be severe when one is coming back up from those depths.

During descent there are four important pressure effects, described below: aerotitis media (or “barotitis”), the “squeeze,” nitrogen narcosis, and high-pressure neuroleptic syndrome. Problems with compression-induced barotrauma arise when there is no free communication into an enclosed space by which to equilibrate pressures; a pressure gradient then forms, with increased external pressure pushing against a negative relative pressure inside the enclosed space.

Aerotitis media is an injury to the middle ear that occurs from pressure across the tympanic membrane when the Eustachian tube is occluded because of allergy or upper respiratory infection. Because the pressure is outside the occluded passage pushing “in,” the tube cannot easily open to equalize pressure, as it does in decompression. The round window and tympanic membrane are easily traumatized and may rupture. Symptoms of otitis may be present, with fullness, pain, diminished hearing, and vertigo. It is controlled with antihistamines and decongestants, by treating the underlying cause, and by performing a Valsalva maneuver while descending. In general, however, it should be prevented by avoiding compression during and for a time after one has experienced allergies or infections involving the upper respiratory tract.

The “squeeze” results from the use of a rigid helmet without adequate pressurization of the diving suit, or when a hole is ripped in the suit. The pressure gradient forces blood from the body toward the relatively negative pressure of the head, resulting in vascular engorgement, dyspnea, petechial hemorrhaging, and epistaxis. This is a problem with supplied-air diving using a helmet, which is now uncommon.

Nitrogen narcosis is a condition occurring when moderately deep dives (below 30 m) are attempted while breathing air. It is caused by the central nervous system effects of dissolved nitrogen at high pressure. There is a prodrome in which the affected diver feels numbness around the mouth. This progresses to euphoria and symptoms

resembling alcohol intoxication, including ataxia. Hallucinations and loss of consciousness may follow. Obviously, these effects are extremely dangerous in diving situations and may lead to incidents resulting in drowning and may also endanger the safety of other divers on the team. Nitrogen narcosis is prevented by using diving gas mixtures with reduced nitrogen content or substituting helium.

The high-pressure neuroleptic syndrome is a condition of central nervous system hyper-excitability seen under conditions of prolonged saturation diving, undertaken at depths far below those at which conventional industrial activity takes place. It begins with deterioration in mental functions and clonic jerks, and progresses to seizures and obtundation.

## **Decompression**

Decompression effects are much more common than compression effects, with the possible exception of mild degrees of aerotitis. They occur when a diver returns to surface too quickly or when workers who are in compressed chambers, such as caissons, have been depressurized too abruptly. All commercial and amateur divers are instructed in the use of diving tables, a set of graphs and tables that provide guidelines for ascent after diving to a given depth and remaining there for a given period of time. Failure to adhere closely to these tables, through inattention or during an emergency, may result in potentially serious effects, including decompression sickness, air emboli, and aseptic necrosis of bone. Recently, the tables have been reevaluated and are thought by some experts to be insufficiently protective.

Decompression sickness is often categorized into three types:

*Type 1:* Pain, urticaria, pruritus (corresponding to a mild case of “the bends”)

*Type 2:* Neurological changes, chest pain, cardiovascular abnormalities (indicating risk of gas embolism)

*Type 3:* Aseptic necrosis of bone (a long-term effect)

Decompression sickness ("the bends") is the result of the out-gassing of nitrogen that has dissolved into body fluids under high pressure. Because nitrogen is relatively insoluble at normal pressures, bubbles of it are not quickly absorbed and carried away from tissues in blood, as are the bubbles of oxygen and carbon dioxide that also form. Instead, the nitrogen bubbles may accumulate and disrupt tissue perfusion by blocking capillaries, causing emboli, and aggregating platelets. Where and when the ischemic effects may occur are highly variable. In its mildest form, decompression illness may consist only of pruritis ("diver's lice") and aching joints. If the bubbles form in the spinal cord or other critical tissues, the condition can cause permanent damage and may even be fatal. Signs of decompression sickness after a dive below 10 m constitute a medical emergency. Treatment is by repressurization, either by immediate return to depth followed by proper decompression procedures, or by repressurization and gradual release of pressure in a surface decompression chamber. Such chambers are available on an emergency basis in most coastal metropolitan areas and at naval facilities in the United States and Canada. Preventive measures include following appropriate diving guidelines and avoiding air travel following a dive, since airline cabins are not fully pressurized and therefore add some degree of further decompression.

Air embolism is a medical emergency. Air emboli result from overly rapid decompression, particularly when a diver fails to keep exhaling on ascent or attempts to hold his or her breath. As the external pressure drops, the relatively increased pressure in the lung causes highly positive intrathoracic pressures, overdistension of air spaces, pneumothorax, mediastinal and interstitial emphysema, and entry of air into the pulmonary venous circulation, where it collects in the left ventricle and embolizes as bubbles. The gas emboli may obstruct perfusion in the cerebral vascular bed, causing a stroke. This condition may be rapidly fatal or may result in serious neurological deficits. Placing the diver in the left Trendelenburg position (on left side, head dependent) directs the emboli away from the brain and increases hydrostatic pressure for perfusion of the head. Recompression and gradual decompression in a

chamber are then required to re-dissolve and then remove the gas emboli through slow diffusion of dissolved gas.

Aseptic necrosis is a result of bubble formation and impaired perfusion in bone. It may occur months or years after repeated episodes of decompression sickness or after an isolated incident of over-rapid decompression.

## **High Altitude**

Effects associated with ascent to high altitude and reduced barometric pressures tend to be complicated by hypoxia, cold, and acceleration during flight. These are abundantly and elegantly dealt with in the literature of aerospace medicine, and are too highly specialized and uncommon in general occupational medicine to describe in detail. This section provides only a brief overview.

Rapid ascent to a high altitude (7000 to 9000 m or more) without pressurization produces an effect much like that of a relatively mild case of decompression sickness. The critical need for survival in such a situation is oxygen. The symptoms of decompression respond to descent to lower altitude or repressurization.

Although commercial airplane cabins are pressurized for safety and passenger comfort, they are not usually fully pressurized to 1 atmosphere, and changes in barometric pressure during flight are easily discernible by any air traveler. Air trapping in the middle ear when the Eustachian tube is blocked or in facial sinuses can cause painful effects. Passengers with serious respiratory disorders may become hypoxic with the reduced atmospheric pressures in an airline cabin and should travel, if they must, with their own supplemental oxygen. Airlines will not usually assume responsibility for providing oxygen to travelers, except as part of emergency procedures.

“High-altitude sickness” is a condition associated with longer periods at higher altitudes without sufficient acclimatization during the trip up. It presents with changes in mental function, personality (including aggression or over-confidence, which can lead to lethal misjudgment), and ataxia, which are the results of “high-altitude

cerebral edema" (HACE), or dyspnea and oxygen desaturation, which are the results of "high-altitude pulmonary edema" (HAPE). Both of these conditions are rapid in their evolution and are potentially lethal unless the patient is brought down to a lower altitude. There is great individual variation in susceptibility to high-altitude sickness and in susceptibility over time. HACE may respond to dexamethasone, and acetazolamide may prevent mild cases of HAPE but is not reliable. Descent is the only reliable treatment. Few industrial activities are carried out at elevations this high, however.

At high altitude, there are many other risks besides atmospheric pressure. Dehydration is a serious risk, as is hypoxemia. Residents of high mountain communities, as in the Andes, develop a compensatory polycythemia to cope with the low oxygen tension. Andean natives also commonly chew coca leaves, both as a stimulant and as folk prophylaxis against high-altitude sickness.

## **TEMPERATURE EXTREMES**

Internal temperature regulation in the presence of temperature variation in the environment is an essential mammalian characteristic and necessary for human life. Problems arise when one of three conditions occurs:

- Temperature variations are so extreme that they exceed the considerable ability of the body to adapt.
- Mechanisms of adaptation, such as vasodilation or sweating, are impaired.
- Exposure to extremes of temperature is concentrated in a particular body part, as in frostbite or thermal burns.

The human body regulates temperature through the central nervous system from a control center in the hypothalamus. This center receives neural impulses from thermal receptors on the skin and from receptors sensing the temperature of blood in deep body structures. It responds

by activating mechanisms controlled by the autonomic nervous system that dissipate heat (vasodilation and sweating) or that increase the internal generation of heat (shivering) and conserve heat (peripheral vasoconstriction). It also sends signals to the cortex that make one aware of being hot or cold, and that initiate behavioral changes, such as changing dress, seeking shelter, or modifying activity. This center can be “fooled” into inappropriate responses by infection (usually fever but rarely hypothermia), autonomic dysfunction, alcohol, or the delirium brought on by temperature extremes outside the adaptive range.

The body regulates average temperature in the deep body “core” within a narrow range centered on about 37°C as measured by an oral thermometer. This core temperature is most reliably measured by using a rectal thermometer (unless the lower extremities are unusually cold or the subject is actively exercising) or by measuring the temperature in 100 ml or more of freshly voided urine. Core temperature averages closer to 37.5°C. For most purposes, oral temperatures are sufficient in ambulatory care, but core temperature measurements are necessary in cases of hypo- or hyperthermia for proper clinical management. Human core temperatures show a diurnal variation, peaking at 0.5–1.0°C higher in the evening; and in women these temperatures show variation with the menstrual cycle, peaking at 0.1–0.4°C higher in the post-ovulatory phase. Skin temperature varies widely, as would be expected given that cutaneous vasodilation is a mechanism of homeostasis.

Although core temperature is maintained approximately constant, there is continuous variation in the heat flux needed to maintain this constancy in the body. Heat is generated by metabolic processes and by work performed by the muscles. Heat is also taken into the body from the environment if the external temperature is warmer than the body core. Heat is lost to the environment by four means: radiation from the surface of skin (as infrared radiation), evaporation in the form of sweating, conduction by contact with a cooler surface, and convection by air movement carrying heated air away from the surface of the skin or expired air from the lungs. Expired air from the lungs is saturated with humidity and is therefore able to carry much

more heat than dry air; it is therefore an important mechanism of heat loss through evaporation and convection. Heat cannot be lost as efficiently from the body when there is interference with these mechanisms. Radiation and conduction may be reduced by insulation, as with padded clothing, or evaporation and convection by restricting air circulation close to the skin. Evaporation is also reduced when humidity is elevated.

In contrast to the homeostatic regulation of temperature, skin temperature varies considerably and registers as the perception of temperature. Perception of heat and cold is highly subjective and a matter of individual preference. Current norms and standards for heat and humidity are based on comfort for the largest proportion of workers and may be perceived as uncomfortable for a large minority.

### **Cold Stress**

Fewer physiological mechanisms exist to conserve or to generate heat in the body than exist for cooling. The principal physiological mechanism to increase heat production is shivering. The major mechanism to reduce loss of heat from the skin is vasoconstriction. Neither mechanism is particularly effective or can be sustained for long periods. As a consequence, human beings are dependent on behavioral mechanisms to keep warm. Clothing, heating technology, wind screening, increasing exercise levels, and assuming positions that conserve heat (such as huddling with others or curling up tightly) are the principal means that human beings have to stay warm once their limited adaptive capacity is exceeded. The primary risk of generalized cooling is hypothermia, a rare occupational health problem even in extreme Arctic or alpine environments. More common are specific cold-related disorders affecting exposed body parts.

### ***Cold-Related Disorders***

Chilblains (perniosis) and minor degrees of skin freezing, called "frost nips," are common but result in no permanent injury. In

occupational settings, serious problems with cold fall into the following categories:

*Frostbite and freezing of exposed body parts* (often the ears, nose, toes, and fingers) usually occurs because clothing does not adequately cover the face or hands, or because it becomes wet from perspiration and conducts heat away from the body. Constriction of blood flow by overly tight boots or elastic sleeve bands predisposes one to frostbite, as does dehydration, smoking- or drug-induced vasoconstriction, or hypoxia, usually due to high altitude. Contact with conducting surfaces, such as metal, may cause frostbite. Once near freezing, the skin becomes numb and the affected person often does not realize that frostbite has occurred until after he or she reaches shelter and warms up. Superficial frostbite, on the surface of the skin, is an annoyance and may be painful, but recovery is usually complete after some blistering and discomfort. Prolonged freezing damages cells in deeper tissue and may lead to gangrene. The affected part should be warmed rapidly with warm water no hotter than 45°C until the tips of the affected part are pink or red, but no longer; hot water may compound the damage by scalding the remaining healthy tissue. Obviously, rubbing with snow only adds to the effect of frostbite and should never be done as a first-aid procedure. If the affected part is thawed and then refrozen before complete recovery, massive further tissue destruction results. Therefore, if freezing occurs far from medical facilities in an uncontrolled or wilderness situation, it is better to keep the body part frozen than to risk thawing and refreezing. Tissue swelling after thawing may result in compartment syndromes, and compression of blood vessels may require fasciotomy on an emergency basis. Severe frostbite is a medical emergency requiring surgical debridement and sometimes amputation.

*Trench* (or “immersion”) *foot* is a condition caused by arteriolar constriction induced by cooling of the feet below 20°C but above freezing in wet conditions for several hours. This may cause sensory and motor paralysis, muscle damage, and swelling. There are mottled red and blue splotches on the affected foot, which looks waxy but remains soft. The skin is friable. The affected worker experiences

great pain when the foot is re-warmed and blood circulation returns; this circulation causes the foot to swell and turn red. Treatment includes gentle re-warming, pain medication, aspirin (to reduce the risk of thrombosis), and alcohol (to increase vasodilation). Recovery is slow and may be incomplete, leading to permanent disability. Complications include gangrene and infection. Appropriate, non-constricting footwear is critical for work in cold climates, and footwear and socks must be changed as quickly as possible when they become wet.

*Immersion in cold water* is a serious problem among workers on offshore oil rigs, onboard ships, or involved with diving. Rapid cooling of the body may be fatal in five minutes or so in the cold waters of the north Atlantic or Arctic ocean, for example. Loss of consciousness also may result in drowning or collision with debris. Although children who are immersed in cold water for half an hour or more often survive unimpaired, in large part because of an intact mammalian diving reflex, adults are much more likely to sustain severe brain damage if they survive prolonged immersion. Even so, it is a tenet of hypothermia management that resuscitation efforts be instituted even in apparently hopeless cases, as surprises do occur. (There is a saying in hypothermia management that “they aren’t dead until they’re warm and dead.”) Diving in cold waters is normally done with a highly efficient wetsuit; and drilling rig workers, pilots, and sailors in very cold waters should wear special hypothermia suits designed to insulate the body and keep the user afloat. Treatment is the same as for dry hypothermia.

### ***Hypothermia***

Most hypothermia cases are associated with immersion in cold water, a situation that is a well-known hazard and therefore guarded against in industries where a hazard may occur, such as in offshore oil production in the North Sea or Hibernia fields (Canada).

Hypothermia, whether of the immersion type or dry, is a medical emergency requiring access to critical care facilities. To manage

hypothermia effectively, thermometers capable of registering rectal temperatures well below 32°C are needed, as this is the critical cut-off for clinical management. Key symptoms of early hypothermia include deficits of judgment and short-term memory and ataxia. More profound hypothermia may present with uncontrolled and exaggerated shivering, followed by cessation of shivering, and obtundation. Dehydration is invariably part of the process. Oral administration of warm fluids alone is not sufficient to re-warm conscious hypothermic patients, and may induce cutaneous vasodilation that shunts heat away from the body core, where it is needed most. Rapid warming by peritoneal dialysis is the treatment of choice for severely hypothermic individuals, those with core temperatures below 32°C.

The clinical management of severe hypothermia is beyond the scope of this chapter, but it should be noted that immersion in warm water and a heating blanket is generally effective, with escalation to gastric and bladder lavage and even cardiopulmonary bypass if necessary. Heated and humidified oxygen helps to warm the heart during ventilation. The patient must be monitored closely for arrhythmias and electrolyte imbalance during re-warming; there is a high risk of asystole, associated with poor chances of survival, or ventricular fibrillation, which has a somewhat better prognosis after defibrillation. Intubation frequently precipitates arrhythmia and therefore should be delayed as long as possible.

Hypothermia of the type seen among the urban homeless in winter is rarely a problem in occupational medicine because of the degree of protection normally provided to workers in extreme climates, and the relative fitness of the working population. Risk factors predisposing a person to hypothermia include dehydration, infection, certain drugs (especially barbiturates, antidepressants, phenothiazines, and hypnotics), poor physical conditioning, physical exhaustion, inadequate nutrition (especially inadequate intake of fats), and alcohol abuse. For employment in Arctic work camps, employers typically select their workers stringently, provide lavish high-calorie meals, and ban alcohol and drugs entirely. The risk of hypothermia can be considerably reduced even further by active physical conditioning prior

to working in cold climates, adequate rest periods, and good nutrition and fluid intake. Protective clothing is a mainstay in the prevention of hypothermia and cold-related injury.

### ***Protective Clothing in Cold Weather***

The selection of proper protective clothing is critical to comfort and safety for work in cold weather. The best insulating material available is air, and an important part of good clothing strategy is to maximize the amount of air contained in fabric and in pockets between the skin and the environment. Another important precaution is that water vapor be able to pass as freely as possible through the clothing to prevent the dampness of perspiration from compromising the insulating capacity of the clothing.

The outer garment should be resistant to the conduction of heat, a property measured using a unit called the “clo,” which is similar to the “R factor” used in building insulation. The clo is calculated from an equation that takes into consideration how long the clothing can maintain a temperature gradient between its two sides. One clo is the unit of thermal resistance of clothing that would keep an average human being comfortable while seated quietly in a room at 21°C for long periods—roughly what would be provided by a medium-weight business suit. If the clo values of various garments are known, they can be added to determine the total insulation value of the outfit. The clo value is greatest for materials that trap air but do not readily compress at pressure points, as compression reduces air trapping. Constructions with baffles and flaps to prevent cold air from coming through zippers, and drawstrings to prevent the “bellows effect” from expelling warm air during activity add greatly to the clo value of cold-weather clothing.

The best strategy is to “layer” clothing, so that pieces may be put on or taken off one at a time to accommodate changes in temperature and to allow personal comfort. The trunk and head should be particularly well insulated. The clothing should be loose, with no constrictions to reduce peripheral blood flow at the wrists or ankles. The

material of inner garments should be an open weave of wool to permit air trapping for insulation and to prevent accumulation of perspiration. Synthetic materials should not be used next to the skin because they trap moisture. Cotton has very poor insulating properties and can be dangerous in cold weather; when cotton becomes wet it conducts heat very efficiently and it dries slowly. Cotton-synthetic mixes are acceptable for underwear and inner garments because they dry much faster and insulate better than pure cotton. Down is an excellent insulating material when dry but not when wet or compressed.

In recent years, a number of synthetic materials with microscopic porosities have been developed under various commercial names. These fabrics retain heat efficiently but allow water vapor to diffuse out. They are particularly suitable for outerwear and for wilderness gear such as sleeping bags and tents.

Protection of the face and head is a particular problem. The head is the site of greatest heat loss (proportionate to its surface area) and must be covered with an insulating cap. In extreme temperatures, an attached hood with an air pocket is needed for protection of the face. This is created by extending the hood forward, creating a deep space in front of the face in which warm expired air buffers the intrusion of cold outside air. This transition zone protects the face from frostbite and reduces heat loss, although it obstructs peripheral vision.

Care should be taken not to dress too warmly. A common mistake is to wear too much or too heavy protective clothing in cold weather. For work in cold climates it is usually more comfortable to be slightly cool, not at "room temperature." Clothing should allow the heat generated during work to dissipate in order to prevent overheating.

The most convenient and versatile cold-weather clothing for most purposes is a parka based on the design of the Inuit peoples of the Arctic. These light garments are remarkable for their insulating ability and are traditionally supplemented by the Inuit with fur overgarments and overpants when the temperature drops to extremes (such as  $-40^{\circ}\text{C}$  and below). Modern parkas are constructed with the traditional features of hoods and double layers of insulation and are very

comfortable in extremely cold conditions. Jeans and other denim garments are particularly poor choices for cold weather, as they provide virtually no insulation, consist of cotton, and are tight-fitting.

### **Heat Stress**

Heat stress is not limited to tropical climates or jobs involving proximity to a heat source. It can also occur as the result of excessive heat retention due to the combination of heavy clothing and strenuous exercise, or in combinations of heat and humidity that interfere with evaporative cooling.

To maintain the body at optimal temperature, it must be cooled at a rate comparable to the heat it generates, in direct proportion to metabolic work. Heat is lost from the body by convection when it is carried away by cooler air, by radiation from the body into cooler surroundings, and by evaporation of sweat from the surface of the skin. Evaporation or artificial cooling must balance the heat gained from convection and radiation when the surroundings are hotter than the body. If heat loss does not equal heat gain and heat generation, then heat accumulates and the core temperature rises.

There are four well-recognized medical conditions that can develop as a result of heat stress. Of these, heat stroke is the most serious and will be discussed in greatest detail.

*Miliaria rubra*, or “prickly heat,” appears as a maculopapular eruption caused by local heat retention, usually on covered parts of the body. Symptomatic treatment with talcum powder or calamine lotion provides symptomatic relief.

*Heat fatigue* or heat cramps present with symptoms of weakness, fatigue, and muscle cramps and is caused by salt depletion, often aggravated by dehydration. These symptoms improve with rest and replenishment of fluids.

*Heat exhaustion* often affects workers who are not acclimatized to working in hot environments. It usually presents as dizziness, headache, orthostatic hypotension, and fainting. The symptoms include headaches, weakness, dizziness, cold and clammy skin, profuse

sweating, vomiting, blurred vision, and epigastric discomfort. This disorder can resemble the early stages of heat stroke, but in heat exhaustion the patient retains normal core body temperature (usually lower than  $38.3^{\circ}\text{C}$ ) and retains a clear sensorium. Serum CPK may be somewhat elevated. This condition results from salt depletion, dehydration, hypovolemia, and vasodilation, leading to cardiovascular insufficiency in addition to the heat stress. Treatment involves oral rehydration, cooling, and rest. Recovery usually occurs within a few hours.

*Heat stroke* is a life-threatening condition resulting from a failure of the thermoregulatory apparatus. It is characterized by nausea, vomiting, hyperthermia, headache, dizziness, and altered mental status, including confusion, delirium, convulsions, and coma. Coagulopathy, causing multiple organ system failure, acute renal failure, rhabdomyolysis, and cardiovascular collapse, is the usual cause of death. The patient usually feels thirsty but may not be able to communicate this while in a confused state. The patient may not perceive heat and may actually feel cold while the condition is evolving. The skin is usually hot and dry, but the absence of perspiration should not be relied upon for the diagnosis. The core temperature rises to  $>41^{\circ}\text{C}$ .

Heat stroke is a medical emergency. One should not be hesitant to make a diagnosis of heat stroke in an appropriate situation. Death can occur rapidly from either high-output cardiac failure, myocardial infarction, or cerebral edema. Although resuscitation may be successful initially, numerous intercurrent complications may arise. These include acute renal failure (10 percent overall risk) related to hypotension, disseminated intravascular coagulation, or rhabdomyolysis, and electrolyte disturbances, such as hypocalcemia, hyperphosphatemia, hypokalemia, or hyperkalemia. Management in its early stages includes removing restrictive clothing, cooling with cold water or water-soaked towels, maintenance of an airway, giving fluids by mouth (if possible) or intravenously, and immediate transfer of the patient to a hospital. The patient should be maintained in a cool, well-ventilated environment. Definitive treatment is effective, rapid

cooling until the core temperature is below 39°C, after which active cooling should be stopped to allow normal physiological mechanisms to take over. Spraying the body with water droplets at about 32°C is the most effective means of cooling. Hydration, restoration of electrolyte balance, respiratory support, treatment of coma and convulsions, dialysis in renal failure, and monitoring for arrhythmias are mainstays of supportive treatment. Intensive supportive care may be required for prolonged periods. Heat stroke has a 50 percent mortality rate despite vigorous therapy.

Deaths from heat stroke are also sometimes reported in otherwise healthy young people engaged in moderate to heavy physical activity, usually in the context of strenuous work outdoors or near a heat source, during military exercises, or at athletic practice. Ambient temperature and humidity are not always extreme in such cases. Most heat-related deaths occur during heat waves in people with an underlying disease, usually cardiovascular, whose condition is decompensated by the additional heat load. Certain conditions predispose one to heat stroke, including obesity, alcoholism, use of certain drugs (anticholinergics and major tranquilizers), acute or chronic illness, fatigue, overeating, poor physical condition, lack of sleep, and older age. The prognosis in heat stroke is greatly improved if the symptoms are recognized early and emergency measures are instituted immediately.

### ***Prevention of Heat Stress***

Heat stress can be prevented by reducing heat gain, increasing heat loss, and acclimatizing the worker. Acclimatization requires a period of adjustment to the temperature, limiting the level and duration of activity until the worker is adapted. Clothing worn in hot workplaces must not be restrictive or occlusive to evaporation and should be as light as possible.

Reducing heat gain can be accomplished by removing, containing, or isolating the source of the heat in the workplace and by installing effective ventilation. Light clothing reduces radiant heat against the skin and removes some heat convection. Maximal exercise

should be kept to short intervals in hot environments and frequent rest breaks may be necessary.

Unacclimatized workers lose sodium in sweat; acclimatized workers retain the sodium and do not require supplementation. Osmotically balanced drinks serve as supplementation without risk, but salt tablets are contraindicated. Salt tablets provoke nausea in some people and may cause stomach cramps. Slow but sustained acclimatization before full work performance is more effective in preventing heat exhaustion than is salt administration. Salt should be added to the diet, however, especially during prolonged work in hot environments.

Effective prevention beyond controlling exposure to heat is largely a matter of ensuring adequate hydration by making lukewarm water freely available at the workplace and by avoiding alcohol or cold drinks. Alcohol dilates blood vessels and can cause fainting. Cold drinks may cause stomach cramps in excessive heat. Drugs that dilate blood vessels or that interfere with autonomic function and sweating are dangerous in hot settings. Alcohol abuse is an important risk factor for serious heat stress. Workers should be encouraged to drink as often and as much as necessary in excessively hot environments, and restroom facilities should be readily accessible to prevent any reluctance on their part to do so. Special heat-resistant protective suits, including air- or water-cooling systems and built-in thermocouples to control temperature, have been designed for unavoidable or emergency work in extremely hot environments. Frequent rest breaks in cooler surroundings help by allowing the accumulated heat to be dissipated, thereby keeping the workers' core temperature near normal.

ACGIH recommendations allow for exposure to heat stress at a level to which nearly all workers may be repeatedly exposed without adverse health effects, assuming that they are acclimatized, fully clothed, and have adequate water and salt intake. A series of curves relating the humidity and temperature yields an index known as the wet-bulb globe temperature index, which is associated with a recommended work/rest schedule. There is no existing OSHA standard, but several individual standards allude to measures for heat stress

prevention and provision of a safe working environment, and protection from heat is covered by the “general duty clause.” (See Chapter 28.) The U.S. National Institute for Occupational Safety and Health (NIOSH) has recommended specific preventive measures to decrease the risk of fatal heat stroke:

1. Acclimatize new workers and workers returning from vacation or an absence due to illness. This can be accomplished by progressively increasing heat exposure and physical exertion over a two-week period.
2. Implement a work/rest regime, matching the severity of heat exposure.
3. Schedule hot operations during cooler parts of the day.
4. Make water readily available at the worksite.
5. Wear protective clothing (i.e., loose, open-weave, light-colored clothing to reflect the heat).
6. Monitor the temperature at the worksite.
7. Perform pre-placement evaluations and periodic health evaluations to identify workers at increased risk.
8. Avoid exposure of members of high-risk groups to jobs with a high potential for heat stress, including alcohol abusers, the obese, and the chronically ill.
9. Instruct workers and supervisors on preventive and emergency measures.

### ***Heat Syncope***

Heat syncope is an episode of transient loss of consciousness associated with heat, posture, and sometimes moderate dehydration. It is caused by essentially the same mechanism as postural hypotension, with heat serving to exacerbate vasodilation.

Heat has a particularly pronounced effect on the circulatory system, acting in several ways to induce a lowering in blood pressure.

These effects include dilation (opening up) of the blood vessels in the skin and surface of the body in an effort to shunt blood (which is warm) closer to the surface so that it can be dissipated, cooling the core of the body. Blood also tends to collect in the dilated vessels of the lower extremities, particularly when one is sitting or standing. This becomes a potentially serious problem under conditions of heat stress because the heart depends on the return of blood to prime itself as a pump; if a reduced volume of venous return is delivered to the right heart, the heart pumps less efficiently. Under normal circumstances, the heart compensates by pumping more quickly, making up for its inability to move a large enough volume of blood with each stroke by increasing the number of strokes in the same time period. Under conditions of heat, however, the cardiac output is already at, or close to, maximum and cannot be increased further; additionally, the heart may be functioning relatively inefficiently compared to its usual work capacity, especially if there is dehydration.

Common situations in which these factors result in problems include fainting spells during hot days among military recruits, the elderly, and occasionally children; occupational exposures to heat without adequate protection; and the lightheadedness one feels after jumping from a lying to a standing position abruptly. These effects are greatly increased with exercise, certain cardiovascular drugs, conditions of dehydration, consumption of alcohol, and old age. Exercise itself causes a redistribution of blood, shunting it away from the central circulation and into exercising muscles.

“Heat syncope” is something of a misnomer, as the mechanism involves a number of physical and hemodynamic factors to which heat is only a contributing cause. Indeed, exposure to heat does not have to be pronounced for syncope to occur with static erect posture or with orthostatic changes in posture from supine to erect. Other causes of vasodilation, which diminishes peripheral vascular resistance, results in blood pooling and limited return to the right heart, and reduced cardiac output, produce the same effect. There are reflexes in the human body that act to maintain blood pressure and prevent syncope, but they can be temporarily overcome with rapid

changes in posture (the reason one feels lightheaded after quickly rising up from bed in the morning) or blunted by drugs, disease, or exercise and physical factors such as heat. Syncope occurs when the blood supply to the brain as a whole is interrupted. This occurs as a result of blood pressure that is insufficient to push blood uphill by the additional 20 cm, more or less, necessary to push the blood to the brain from the level of the heart. Imagine a thirteen-story building served by the city's water supply. The water pressure in the city system is not sufficient to push water higher than the third or fourth floor. An additional pump is required to bring the water up to the level of the thirteenth floor; any failure of this pump to generate the pressure needed will cause the thirteenth floor to go dry. If all the toilets are flushed at once, the pressure will drop and the water supply will also be interrupted for an instant. Likewise, a disorder that interrupts the heart rate, reduces the strength of the heartbeat, or delivers blood away from the vessels carrying it to the brain will cause a drop in pressure, inadequate blood supply to the brain, and syncope.

When syncope occurs, the person affected falls down or slumps over unless there is something to catch him or her. Once the victim is on the ground or in a position with the head below the level of the heart, the syncopal episode disappears. This is because the pressure difference needed to push blood to the brain from the heart in an upright position is no longer necessary, and blood is supplied to the brain. The affected person usually wakes up and recovers completely, unless there has been trauma from the fall. If he or she does not recover completely, other problems should be suspected and investigated.

## VIBRATION

Whole-body vibration is common in industry and in daily life. It has been recognized in recent years as a problem of some concern when it is prolonged and excessive. Segmental vibration is a more limited problem but is also widespread in industry.

Vibration occurs in three dimensions and is described by the frequency (in Hz) and the acceleration, expressed as root-mean-square velocity (in m/sec) in each of the three planes in space.

### **Whole-Body Vibration**

In this context, whole-body vibration is defined as periodic displacement of the worker in three dimensions (primarily vertical) at or near a frequency of 3 to 7 Hz at an acceleration of 0.03 to 0.3 G. (In this context G indicates gravitational acceleration.) At about 5 Hz the human body in a seated position in a moving vehicle absorbs the most mechanical energy because the trunk is in resonance with the basic vibration of the vehicle.

Low-frequency, whole-body vibration is the most characteristic exposure of vehicle drivers and heavy equipment operators. The spectral characteristics of coarse vibration in large vehicles are surprisingly similar; buses typically vibrate between 0 and 15 Hz at a mean acceleration of 0.1 G.

A 1974 NIOSH study of 1,448 male motor coach operators was designed specifically to investigate the health consequences of this exposure. The NIOSH study identified several disorders as significantly excessive among drivers. These disorders were grouped into three possible categories by mechanism of etiology:

- Possibly associated with vibration per se (disorders of bones and joints; disorders of stomach and upper gastrointestinal tract; and ear, nose, and throat surgery), which would also include the observed increase in low back pain
- Possibly associated with increased intra-abdominal pressure and prolonged sitting (disorders of veins; disorders of gastrointestinal tract generally; abdominal hernia; and surgery for abdominal, proctologic, and urologic disorders)
- More likely associated with social factors relating to the job (injuries and violence, increased rate of influenza)

Truck drivers also showed an excessive rate of hypertension, diseases of the oral cavity, and peptic ulcer. Additional, later studies have demonstrated similar findings by similar techniques, including a follow-up study of long-haul truck drivers by NIOSH in 1976.

Intra-abdominal pressure is readily transmitted through the venous system and hollow organs of the gastrointestinal and urinary tracts. Increased intra-abdominal pressure is part of the involuntary process of response to vibration, bracing the trunk. This, and hydrostatic pressure in a prolonged seated position, tend to exacerbate the development of hemorrhoids, varicose veins, and inguinal hernias regardless of the presence of vibration.

Experimental studies of passive whole-body vibration in humans and primates suggest vascular and gastrointestinal effects. Mean arterial blood pressure, heart rate, cardiac output, oxygen consumption, and minute ventilation all increase with exposure to whole-body vibration in the near-resonance range at accelerations similar to those observed in buses. Performance on a number of mental tasks declines with exposure, but can be partially maintained by periodic rest intervals. Although few studies of gastrointestinal biomechanics during vibration have been reported, a pertinent finding is that colon and stomach pressure increases markedly with vibration. Furthermore, stomach pressure and colon pressure appear to oscillate in such a way as to suggest that the abdominal contents are displaced up and down in the peritoneal cavity as if they were a single mass, a vivid literal description of the idiom "churning stomach."

### **Segmental Vibration**

Segmental vibration presents a particularly difficult hazard that may result in more severe health effects. The degree of hazard depends on the frequency spectrum and the magnitude of the vibration produced.

Segmental vibration is a particular problem with hand-held power tools, including chainsaws, riveting guns, pneumatic drills, grinders, air hammers, chippers, and cutting devices for concrete.

The condition is therefore found in occupations in which these tools are used, particularly in the forestry, construction, and sheet metal industries.

The most serious manifestation of segmental vibration of the hand is “vibration vasculitis,” or “white finger disease.” This is a form of Raynaud’s phenomenon characteristically manifested as spontaneous or cold-induced vasospasm and hypesthesia to fine touch, vibration, temperature, and pain. Vibration-induced vasculitis occurs after exposure to vibration in the 125–300 Hz range, usually after 8,000 hours or more, although it can occur after less time.

The sensitivity to vibration appears to be the first modality to be affected and is the basis for some automated screening tests for early disease. After this first stage, the disorder passes through a stage of worsening neuropathy and autonomic dysfunction, then through a phase of sporadic vasomotor ischemia induced by cold and swelling of the fingers in which the typical findings of Raynaud’s syndrome are apparent. This is followed by a final stage in which the vasomotor changes in the hand become more typical and persist regardless of temperature, and other vascular beds may be affected by transient vasospasm. Bone resorption in the distal phalanges may be visible on films of the hand. The degree of impairment and rate of progression of this disorder can be staged by one of several staging systems in the literature that are based on the extent of the functional changes and the natural history of the disorder.

Inexpensive, convenient screening tests for vibration vasculitis have been developed, and are generally based on early loss of vibration sense and fine touch. The differential diagnosis of this condition includes the many other causes of Raynaud’s disease, including systemic sclerosis, and recovery from frostbite.

Prevention of vibration white finger is best accomplished by selecting machine tools that have dampened vibration spectra through engineering controls. In recent years manufacturers have developed tools that dampen vibration in the troublesome frequency range, but there are many tools on the market and in service that still present a hazard. Maintenance of equipment in good condition may reduce

vibration caused by wear and loose parts. Wearing gloves helps, particularly if there is substantial padding and the vibration frequency is greater than 500 Hz. Grasping the tool as loosely as possible while still retaining control, reducing to a minimum the time spent exposed to vibration, and controlling other risk factors for vasculitis, such as smoking and exposure to cold, are useful. Standards for exposure to vibration specify a variable allowable duration of exposure given a certain level of vibration.

## **RESOURCES**

- Chengalur S, Rodgers SH, Bernard TE for the Eastman Kodak Company. *Kodak's Ergonomic Design for People at Work*. Rochester, NY: John Wiley; 2003.
- Collins LR, Schneid TD. *Physical Hazards of the Workplace*. Boca Raton, FL: CRC Press; 2001.
- McKeown C. *Office Ergonomics: Practical Applications*. Boca Raton, FL: CRC Press; 2008.
- Wald PH, Stave GM. *Physical and Biological Hazards of the Workplace*. New York: John Wiley; 2002.

## **NOTEWORTHY READINGS**

- Baetjer AM. Role of environmental temperature and humidity in susceptibility to disease. *Arch Environ Health* 1968; 16:565–570. [Classic]
- Occupational Health Clinics for Ontario Workers. *Office Ergonomics Handbook*. 4th ed. Don Mills, ON: Occupational Health Clinics for Ontario Workers; 2007. [An excellent, convenient resource, suitable for distribution to patients.]
- Pelmeir P, Wasserman DE. *Hand-Arm Vibration: A Comprehensive Guide for Occupational Health Professionals*. Beverley Farms, MA: OEM Press; 1998.

# 10 COMMON CHEMICAL HAZARDS

Chapter 2 presented an overview of toxicology as it applies to occupational and environmental hazards. This chapter discusses selected chemicals and classes of chemicals in greater detail. No occupational and environmental medicine (OEM) practitioner can know (and remember) all of the chemical hazards that he or she is likely to encounter over a busy practice. Those selected for this chapter are common, important in practice, and part of the daily discourse of the profession. Understanding them in some detail also provides a mental framework or model for understanding similar hazards.

## ALCOHOLS AND GLYCOLS

Alcohols are defined as compounds in which a hydroxyl group,  $-\text{OH}$ , is attached to a saturated carbon atom, represented as  $\text{RH}_2\text{COH}$ . These compounds are widely used in industry as solvents, antifreeze agents, and reagents. Alcohols depress the central nervous system (CNS). Alcohols with multiple hydroxyl groups are called polyols, and the most important of them encountered in practice are the glycols (diols). Glycols are organic molecules bearing two alcohol groups, and are used as heat exchange agents, antifreeze

agents, solvents, and reagents. The glycols are more viscous and less volatile.

Of all alcohols, ethanol, also referred to as ethyl alcohol, is by far the most familiar and the most widespread as a potential exposure. Ethanol is a very helpful mental model for understanding solvent toxicity because the acute and chronic effects are similar. Surreptitious drinking in the workplace, alcohol abuse outside of working hours, and the connection between alcohol abuse and more broadly dysfunctional lifestyles constitute a tragic burden of personal and family injury, in addition to contributing to occupational injuries. Circumstances of ready availability and peer encouragement may result in a worse pattern of alcohol abuse than might have otherwise been the case. The role of employee assistance programs in dealing with alcohol abuse is outlined in Chapter 19. Because of its familiarity to all practicing physicians, ethanol will not be discussed.

The simplest alcohol, chemically, is methanol ( $\text{CH}_3\text{OH}$ ), also called methyl alcohol or “wood alcohol.” It is used as an antifreeze agent and contains particularly desirable properties because it dissolves and removes water that has been trapped as condensed moisture in gasoline lines or frozen as ice. Methanol is also used as model airplane fuel, windshield washer fluid, solid cooking fuel, and photocopying fluid. Additionally, it is a constituent of some paint removers, and in liquid chromatography it is used as a solvent vehicle. Methanol is rapidly absorbed across the skin and, being highly volatile, is easily inhaled and may be absorbed transdermally. Ingestion of methanol is rare in occupational situations, but can occur accidentally or among persons consuming illicit alcoholic beverages from questionable sources. Like ethanol, methanol is capable of inducing inebriation and CNS depression.

Methanol is metabolized by two pathways, including the alcohol dehydrogenase pathway, which it shares with ethanol. (Metabolism of ethanol and methanol is discussed in Chapter 2.) It is rapidly and completely transformed to formic acid, a highly potent neurotoxin causing retinal injury due to edema of the optic disc and axonal damage to the optic nerve. Methanol can cause papilledema and

visual changes. A profound metabolic acidosis occurs with ingestion of large amounts of methanol. The treatment of this medical emergency is to block alcohol dehydrogenase activity by infusing ethanol (the traditional intervention), or fomepizole or 4-methylpyrazole for competitive inhibition, in order to prevent the formation of formic acid. Chronic exposure is much less a problem with methanol than acute, high-level exposure, and no syndrome of chronic toxicity has been identified. Animal studies of methanol are difficult because metabolic pathways are different in non-primate species.

Glycols taste sweet, and the attraction occasionally results in inadvertent ingestion by children, animals, or impaired adults. These alcohols are not easily absorbed transdermally or freely inhaled, although some intake by inhalation may be possible if they are heated. Ethylene glycol, or 1,2-ethanediol ( $\text{CH}_2\text{OHCHCH}_2\text{OH}$ ), is the most common glycol. In the body, it is metabolized to oxalic acid and glycolic acid, causing severe metabolic acidosis and the risk of renal failure. Treatment is by infusion of ethanol or fomepizole for substrate competition for alcohol dehydrogenase. Diethylene glycol ( $\text{CH}_2\text{OHCHOHCH}_2\text{OH}$ ) is also used in cosmetics and in lubricants as a softening agent and plasticizer. Ingested, diethylene glycol is an extremely toxic nephrotoxin, but exposure by this or any other route is rare as the volatility is very low. Propylene glycol or 1,2-propanediol ( $\text{CH}_3\text{CHOHCH}_2\text{OH}$ ) is also used as a solvent in foods and pharmaceuticals, and is an environmentally sound alternative to antifreeze containing ethylene glycol. It is metabolized to lactic acid in the body and appears to have very low toxicity.

Glycol ethers are a class of compounds utilized as solvents, especially in lacquers, printing inks, textile dyes, leather finishing preparations, epoxy resin coatings, and varnish removers, as well as in antifreezes and gasoline additives, often under the general name Cellusolve®. They are heavily used in antifreeze, in hydraulic brake fluids, and in the inks used in stamp pads, ballpoint pens, and print shops. Two members of the class (2-methoxyethanol, or ethylene glycol monomethyl ether (EGME), and 2-ethoxyethanol, or ethylene glycol monoethyl ether, EGEE) have been shown to cause fetal toxicity and

malformations in animal studies, as well as testicular atrophy and male infertility. It has been demonstrated that males exposed to EGME have decreased red blood cell counts and thus higher levels of anemia. In females, associations have been found between EGME exposure and congenital malformations, spontaneous abortions, and sub-fertility. Neurotoxicity has been reported in animal studies and case studies, the latter involving acute encephalopathic symptoms and persistent weakness, which have been associated with transcutaneous exposure to 2-methoxyethanol. Other symptoms that have been reported are personality change, staggering, and marrow effects; these may also occur with other members of the family of compounds, including 2-butoxyethanol.

## **ALUMINUM**

Aluminum is a very common substance in metal fabrication and is a ubiquitous element in the Earth's crust. There are abundant opportunities for workers to be exposed to it both on and off the job, and in the form of oxides and aluminum silicates it is a natural background element in the environment. Aluminum is generally treated as nontoxic because it is rarely encountered in a bioavailable form. Periodically, however, its potential toxicity becomes an issue, especially with respect to neurotoxicity and dementia.

There have been recurrent scientific reports that exposure to bio-available forms of aluminum may be associated with the risk of dementia. To date this is an unproven hypothesis, not supported by the weight of evidence; no studies of workers handling the metal in normal industrial processes have shown any adverse effect. Periodically, however, media stories reappear attempting to link aluminum with the risk of dementia or lung disease, or raising concerns about the use of alum in agglutinating solids in drinking water treatment or cooking with aluminum pans.

For years, investigators have known that aluminum is found in higher concentrations in the brains of persons with Alzheimer's disease, but this seemed likely to be a secondary phenomenon of nonspecific

accumulation by abnormal tissue, rather than a cause of the condition. It is localized to the neurofibrillary tangles that are the hallmark of Alzheimer's disease. Experimental studies feeding animals aluminum did not produce the lesion, however. Dialysis dementia is more clearly associated with accumulation of aluminum.

Without effective absorption and distribution, the metal is unlikely to reach nerve tissue; even if it gets there, an abnormality may be required for intracellular uptake and toxicity. The form of aluminum used as a flocculent for water treatment, alum, is not bio-available. It is difficult to lend credence to the hypothesis that drinking water is a factor in CNS problems because the amount of aluminum from that source and migration from cooking utensils is so small compared to its abundance in food and medications.

Few people are exposed to aluminum in a form that is easily inhaled and distributed intracellularly. The fireworks and flare industry is the only example, and there have been case reports suggesting neurotoxicity in association with extremely fine aluminum powder in oil. (It oxidizes too readily in air.) In the 1930s aluminum powder was introduced as a prophylactic agent against silicosis for the protection of miners. As a treatment to prevent silicosis in workers exposed to silica, it seemed to work very well and was continued in some parts of the world through the 1970s. A common practice was to blow the dust into bathhouses, where the miners coming off shift would inhale it as they changed and showered. The treatment was halted because the intrinsic fibrotic potential of the aluminum dust itself was revealed in animal studies, and because of the fear that treated miners might develop Shaver's disease, an alumina-associated pneumoconiosis initially described in workers handling aluminum-containing abrasives. Subsequent studies of miners intentionally exposed to aluminum powder as a prophylactic measure for protection against silicosis have shown changes in psychological testing scores that suggest a neurotoxic effect. Fortunately, dust suppression techniques improved and the attraction of prophylaxis as an intervention for prevention faded. Control of exposure replaced involuntary treatment of the worker, with all of its ethical problems.

## AMMONIA

Ammonia ( $\text{NH}_3$ ) is one of the most common chemicals used in industry and especially agriculture. It is the primary nitrogen source in fertilizer and is heavily used in the production of chemicals, including explosives, refrigerants and cleaning agents. Ammonia, in anhydrous form, is transported and stored under pressure, and incidents that result in exposure usually occur as a result of leaks, ruptured vessels, or transport accidents. Ammonia is a strong base and its alkalinity can lead to chemical burns. As a consumer product, ammonia is a cleaning agent supplied in dilute solutions of 5 percent in water. Ammonia is a highly irritating gas that causes upper airway irritation, skin burns on contact, and irritation of the conjunctiva and iris when the eye is exposed. Ammonia in water forms ammonium hydroxide ( $\text{NH}_4\text{OH}$ ), which results in injury to the bronchial mucosa when inhaled. Most of the injury is in the upper airway, because ammonia is highly soluble and is cleared before it reaches the lower respiratory tract in most situations. At very high concentration, ammonia can penetrate to the alveolar level to cause pulmonary edema, but this is rare (see Chapters 2 and 17). In these rare cases, overwhelming exposure may result in toxic inhalation and pulmonary edema, but the risk of deep lung injury is much lower for ammonia than for less soluble gases. Because ammonia is lighter than air, it does not collect except in confined spaces.

## ASBESTOS AND OTHER FIBROUS SILICATES

Asbestos is a general term for a heterogeneous group of hydrated magnesium silicate minerals that have in common a tendency to separate into fibers. Asbestos is historically the most significant occupational “carcinogen” (actually a class of loosely related carcinogens), and is responsible for a current epidemic of lung cancer, as well as a smaller epidemic of mesothelioma. For some time this epidemic was buried in national statistics due to the larger epidemic of lung cancers caused by cigarette smoking. These epidemics resulted primarily from exposures associated with the expanded use of asbestos after World War II. The epidemic of asbestos-related lung cancer was, and

still is, seen mostly among individuals who also smoke. Asbestos has been widely banned but is still used in some countries in the developing world. The incidence of future asbestos-related disease will primarily reflect exposures of the recent past. Asbestos is therefore important not only as a major hazard but as a lesson of history.

Asbestos exposure is associated with “asbestosis,” a fibrotic lung disease induced by dust (pneumoconiosis), discrete fibrotic changes of the pleura (pleural plaques), generalized or “diffuse” thickening of the pleura, a form of chronic obstructive airway disease, increased risk of bronchogenic carcinoma (interactive or synergistic with cigarette smoking), and mesothelioma, a very aggressive sarcoma of the pleura. The health outcomes of asbestos are discussed in much greater detail in Chapter 17, because of their importance in occupational medicine.

Asbestos is an artificial family of naturally occurring silicates found in mineral deposits in only a few areas of the world, but it has been disseminated worldwide through commerce, particularly in the 1960s and 1970s by industrial application and distribution of products. Table 10.1 describes the major forms of asbestos, their properties of economic value, and the principal commercial uses in the 1970s, at the peak of asbestos consumption in industry. As will be noted, it is the distribution of uses in that decade that is of primary interest today because it defines risk at the beginning of the latency period associated with many cancer cases seen today.

Asbestos fibers consist of silicone, oxygen, and other elements in various combinations and amounts depending on the fiber type being considered. Three varieties of asbestos are most important commercially, and therefore as occupational hazards. Chrysotile, or “white asbestos,” is the form that was most commonly used in North America, and is characterized by long hollow fibers formed by flat sheets rolled like a scroll. Chrysotile is the only common representative of the serpentine family of asbestos. It is subject to slow dissolution in the lung and is thought to be less potent in inducing cancer, either lung cancer or mesothelioma. This is best explained by the fact that chrysotile fibers dissolve over time, and are cleared from the lungs more efficiently than are their amphibole counterparts; however, this

**Table 10.1.** Varieties of Asbestos

Type	Structure	Property	Use in 1970s	Origin	Mineralogy
Chrysotile ("white asbestos")	$Mg_3Si_2O_5(OH)_4$ . Flat sheets rolled into long, hollow, tubular fibers	Flexible, fine, easily woven, very heat resistant	Insulation, all other asbestos uses	Canada (primary producer), USSR, Italy, southern Africa (not South Africa)	Fibrous serpentine
Crocidolite ("blue asbestos")	Si, Fe, Mg, Na Solid, twice the diameter of chrysotile	High tensile strength	Textiles, pipes, cigarette filters, cement reinforcement	South Africa, Bolivia, Australia	Fibrous riebeckite
Amosite ("brown asbestos")	Si, Fe, Mg Solid	Poor flexibility and heat resistance	Cement pipe, tiles	South Africa	Fibrous cummingtonite-grunerite
Anthophyllite (brown)	Si, Ca, Mg Solid	Resistant to acid, brittle (low tensile strength)	Cement, chemical industry, additive in ceramics, paints, floor tile	Finland, USA, Brazil	Fibrous anthophyllite, orthorhombic (polymorphic with cummingtonite)

Actinolite ("green" or "gray asbestos")	Si, Ca, Mg, Fe Solid	Brittle, very resistant to heat and acid	Chemical industry, contaminates vermiculite	Canada, USA	Fibrous actinolite, monoclinic
Tremolite (creamy white to green)	Si, Ca, Mg Solid	Highly heat resistant	Chemical filters, contaminates vermiculite	Italy	Fibrous tremolite (nephrite), monoclinic.

clearance is not complete. There is a minority opinion among occupational lung disease experts that pure chrysotile is not or is only weakly carcinogenic, but this is not generally accepted. Chrysotile was the form of asbestos most heavily used in North America at the peak of the industry.

The remaining asbestos types of commercial importance belong to the amphibole family, characterized by solid fibers. Crocidolite, or “blue asbestos,” was the least commonly used form in asbestos-containing insulation and cement because it has low levels of heat resistance. Several studies have implicated crocidolite, an amphibole fiber of particularly elongated shape that is prone to breaking into smaller fibrils, as being the most potent asbestos carcinogen. Amosite, which is also thought to be relatively potent as a carcinogen, has been almost as extensively used as crocidolite. Tremolite is found as a minor constituent or contaminant in rock and in deposits of other types of asbestos.

The definition of asbestos is arbitrary. The term properly refers to commercially produced fibrous silicates, excluding those that are not in commercial use. As described by the National Research Council, “[Asbestos] refers to well-developed and hair-like long-fibered varieties of certain minerals that satisfy particular industrial needs.” The “Libby mineral” is a transitional mineral resembling tremolite that was found to have contaminated vermiculite deposits in Libby, Montana. The Libby mineral appears to have the same properties and to confer the same health risks as some forms of asbestos. Zeolite and fibrous mordenite, which are very similar and appear to have the same toxic properties in their fibrous form, are not considered to be asbestos, although they behave the same and may be more potent. (A regulatory gap could easily be filled by simply declaring them to be asbestos, which would be no more arbitrary than the current definition.) Talc is also a fibrous silicate that is used commercially, but it is not considered to be asbestos because it is not used for its fibrous qualities; talc may, however, be contaminated with tremolite. Although there is a talc pneumoconiosis, pure talc does not seem to confer the same risk of cancer. (See Chapter 2 for the definition of a fiber.) Table 10.2 summarizes the non-asbestos minerals and related

**Table 10.2.** Non-Asbestos Fibrous Silicates

	Composition	Properties	Commercial Uses	Countries
Zeolites	Variable: there are 36 zeolites, of which only about 6 are fibrous and resemble amphibole or chrysotile asbestos, including erionite, mordenite, wollastonite	Catalytic in organic reactions	Catalytic reactors; fibrous form is not used commercially	Widespread, esp.: USA, Turkey, China, central Asia, Australia
Talc and pyrophyllite	Talc: Si, Mg Pyrophyllite: Si, Mg, Al Short fibers Soapstone (steatite) is talc in bulk.	Absorbs moisture and oils, prevents surfaces from sticking, heat and acid resistant	Cosmetics, additives, nonstick powder, fillers, paint, pharmaceuticals	USA, Canada, Europe, China, India, Australia
Fibrous glass	Extruded and spun fibers of super-cooled liquid (glass)	High tensile strength, resistance to heat and acid	Insulation for heat and sound, composite building material	Manufactured
Synthetic vitreous fibers	Variable composition. Extruded ceramic, silicate fibers	High tensile strength, resistance to heat and acid	Insulation for heat and sound	Manufactured

silicates. Because asbestos is defined by its structure as a fiber, the various types of asbestos also have their non-fibrous counterparts. For example, amosite is the fibrous form of grunerite, which exists independently of the fibrous form.

“Asbestos” is derived from the Greek word for “unquenchable”; its synonym in many languages, “amianthus,” is derived from the Greek word for “undefiled.” These terms accurately reflect how asbestos was used by the ancients; it was used as a cloth or wick that remained unconsumed by fire, and as napkins and shrouds cleansed by placing them in fire. These properties of asbestos—it could be woven and it resisted fire and acid—determined its industrial applications until the twentieth century, when it was also found to have useful mechanical properties such as a high tensile strength. It has been widely used in cement (especially in cement water pipes, where the addition of asbestos adds strength but is not required), flooring materials (especially linoleum prior to the mid-1970s), brake linings, gaskets, and fire-resistant clothing and gloves. Crocidolite asbestos was also used as filter material in cigarette filters (specifically, the Kent brand Micronite filter cigarette), which created a risk of mixed exposure and interactive effects contributing to the risk of lung cancer. It was also used in gas masks, which put asbestos directly into the user’s inhaled air.

The single application involving the greatest exposure to asbestos, however, has been its use as insulation, both in sheet form and as slurry. This slurry was sprayed onto steel girders and construction members, as well as on the interiors of buildings, to prevent structural failure due to fire. In such settings, high concentrations of asbestos fibers were present in the breathing zone of insulators and other tradesmen working nearby, in the absence of personal protection. Likewise, shipyard workers were heavily exposed, particularly during the brief period of intensive shipbuilding and outfitting activity just before and during World War II. Almost all of the insulation material used in North America was chrysotile.

Mining and milling of asbestos exposed a relatively small workforce, in those areas where the asbestos deposits were found. The milling processes were at one time more likely to expose workers than mining,

but improved techniques of dust control have reduced the hazard considerably. Factories using asbestos as a raw material often handled it with less than adequate control, at least until recent years. The production of these asbestos-containing products resulted in exposure of many thousands of workers beyond those in the primary extraction and conversion industries. Today, occupational asbestos exposure is a concern in addressing older workers with past exposures, those in building and facility maintenance, those involved in asbestos abatement operations, and those involved in the renovation or demolition of asbestos-containing structures. Asbestos-related exposure is still a concern in the United States for an estimated 1.3 million workers. The bulk of these exposures are in the construction industry as well as the maintenance of buildings and equipment.

Asbestos was banned for most uses in the United States in the 1970s. Concern over asbestos left in place, especially in schools, led to widespread removal of asbestos insulation in the 1980s and 1990s. This had the unfortunate effect of exposing workers involved in removal operations, since those operations often were not undertaken under the strict controls that were required for such a procedure. OSHA regulations for asbestos abatement are detailed and require monitoring, posting of results, personal protective equipment, barriers (sealing off the work area with plastic sheeting), and worker training. Vacuums used in areas where asbestos is present should be equipped with HEPA filters to prevent fiber dispersion. The reality is that today the greatest opportunity for exposure is in maintenance and asbestos abatement conducted by fly-by-night contractors, usually with young, ill-trained, transient, and often immigrant workers.

Environmental exposure to asbestos and asbestos-like fibrous silicates may occur as a natural process wherever these materials are exposed, as in the Cappadocia region of Turkey (where mesothelioma rates are quite high), and in the U.S. Great Plains (where zeolite deposits are weathered and exposed). The release of asbestos into the air in urban areas is a result of the demolition of buildings, or the removal of asbestos from schools or commercial buildings. Children have been exposed by playing in piles of asbestos-containing waste at

mines or construction sites. Exposure to asbestos has also occurred when workers in trades involving the material have brought it home on their work clothes. Housewives and children have been exposed in this way and have developed disorders related to asbestos. Community exposure of the Libby mineral may also have occurred from deposits and storage of contaminated vermiculite near railroad bulk freight-handling stations and warehouses.

The history of asbestos as an occupational hazard is often used to illustrate corporate resistance to worker protection and denial of occupational risk. Table 10.3 provides a brief chronology of significant events in the history of human use of asbestos. Initially, concern focused on the characteristic pneumoconiosis caused by exposure to asbestos. After the turn of the century, British authorities recognized that asbestos caused the chronic lung disease now called asbestosis and initially acted to reduce exposure levels. In the 1930s, the British discovered the additional association between lung cancer and asbestos. Throughout this period, strong resistance and suppression of information by the asbestos industry held back regulation and delayed legal recognition of the liability of asbestos producers. The use of asbestos dramatically increased during armament in preparation for and then execution of combat in World War II. Concerns regarding health effects were largely shelved because of the immediacy of the war effort. In the 1960s, Irving Selikoff investigated insulation workers in New Jersey and New York, reviving and confirming the earlier studies, and adding many additional observations. These studies led to definitive recognition of the risks associated with asbestos exposure, including asbestosis (the pneumoconiosis), lung cancer, and mesothelioma. Many asbestos producers were sued under “third-party liability” theories involving failure to warn workers, to give adequate notification, and to permit safer alternatives on the market. In a bid to avoid exposure to liability, some of these producers declared bankruptcy based on projected claims at face value, although they were still financially solvent. There followed a long struggle, which is still ongoing, between asbestos-exposed workers and these companies with respect to liability and compensation.

**Table 10.3.** Chronology of Asbestos Uses and Asbestos-Related Diseases

2000 BCE	Used in pottery in Finland and to pack cracks in log cabins.
400 BCE	Used as wicks in Greek temple lamps, and for funeral shrouds used for cremation.
100 BCE	Early references by Pliny to weavers of asbestos cloth and their lung problems.
CE	Used in Europe and the New World as “salamander’s wool” or “salamander cotton,” because salamanders were believed impervious to fire.
1880s	Widespread commercial exploitation began in Quebec; after 1905 growth was exponential, primarily chrysotile use.
1898	H.M. Inspector of factories sounds alarm in Britain. Advises control of exposure based on probable health effects.
1899	Auribault reports increased mortality among asbestos workers in U.K.
1900	First documented case of fibrosis in a London asbestos worker.
1903	First use in brake linings. Isolated references to asbestos-related disease in years following, especially by Montague Murray.
1918	Prudential Insurance Company refuses to issue policies for asbestos workers.
1919	15 U.S. asbestos workers with abnormal chest x-rays reported by Pancoast, a radiologist.
1927	Cooke described interstitial fibrosis and coined the term “asbestosis.”
1927	Metropolitan Life Insurance Company performs first U.S. study.
1935	Lynch and Smith reported a case of bronchogenic cancer in an asbestos worker (lung cancer unusual at that time). Thereafter a large body of literature developed.
1955	Series of epidemiologic studies, beginning with Sir Richard Doll, establishes elevated risk of lung cancer.
1960	First report (S. Africa: Wagner) of association with mesothelioma. Later confirmed by Newhouse and others.
1964	First public presentation of findings of studies in insulation workers by Irving Selikoff and Jacob Churg. Cancer risk definitively established.

By the 1980s, asbestos was variably estimated to be responsible for about 1–4 percent of cancers, especially in areas such as coastal locations where shipbuilding labor used asbestos in large quantities 40–50 years ago, and around asbestos mining communities, principally in Quebec. As of today, the United States is still grappling with an epidemic of asbestos-related cancer resulting from exposures occurring around the time of World War II and thereafter, when the use of asbestos greatly increased. This epidemic of lung cancer was somewhat hidden within the larger epidemic due to cigarette smoking, but the influence of asbestos exposure became clearly visible with the startling increase in cases of pleural mesothelioma, previously rare. This pattern of asbestos-related disease incidence is currently being repeated in developing countries, where the peak in asbestos use came later than in Europe and North America.

Lung cancer associated with occupational exposure to carcinogens is not rare, although smoking remains the greatest cause of most lung cancer. Historically, asbestos has been the single most important occupational pulmonary carcinogen, but there are a number of others. Cigarette smoking and asbestos exposure are interactive in risk for bronchogenic carcinoma.

### **Exposure Assessment**

In evaluating a possible occupational association with a patient's malignancy, a complete and accurate history of the patient's employment, as well as the specific jobs performed, is critical. The latency period between exposure and detection of the malignancy is usually on the order of two or three decades. Relatively brief employment decades ago may be forgotten, but may be highly significant.

In the present day, a colorimetric test is available in kit form to identify the presence of bulk asbestos in suspect material. It is not quantitative, however.

Phase contrast microscopy (PCM) has been the standard practice for asbestos determination and monitoring, and is adequate for most purposes. In general, PCM is preferred for routine work within the

level of sensitivity of the technique. Electron microscopy is useful in research or in monitoring extremely low levels. A guiding principle is that a technique and a laboratory, once chosen, should be used throughout, because the absolute number of fibers reported or the estimate of mass is much less important than achieving valid comparisons among sites and over time.

Phase contrast (particle counting by membrane filter) uses a conventional microscope equipped with a phase contrast lighting and filter system to enhance visibility of the fibers. Fibers trapped on a filter are counted by a technician; the technician uses the criteria of size and dimension in an attempt to distinguish asbestos from other fibers. Results are reported in terms of fibers per cubic centimeter after a correction is made for the volume of air passing through the filter during sampling. Phase contrast microscopy is relatively inexpensive but requires special equipment and a trained technician. There are two main disadvantages to this technique, however. First, it is not specific for asbestos fibers; thus other, non-asbestos fibers may be included. Second, the smallest visible fibers in PCM are about  $0.2\text{ }\mu\text{m}$  in diameter, while the finest asbestos fibers may be as small as  $0.02\text{ }\mu\text{m}$  in diameter. Thus, the final count should be considered an *index* of fiber concentration rather than a meaningful absolute number. The detection limit for accurate counting by optical microscopy is about  $0.1$  fibers per cubic centimeter (f/cc). Since it is well known that there is a dispersion of small fibers below the accurate detection limit of  $0.1$  f/cc, an absolute count is not necessary as long as this dispersion is reliably similar from site to site. Empirical studies confirm that this is a fair assumption for buildings but not within the human body, where fibers are broken down by cellular activity. PCM is recommended by OSHA, and thus is the technique utilized in virtually all occupational monitoring situations.

Electron microscopy is much more sensitive but more difficult to interpret. Smaller fibers are visible (ten times as many fibers  $>5\text{ }\mu\text{m}$  length compared with optical microscopy), and the results are usually reported in terms of mass ( $\text{ng}/\text{m}^3$ ) rather than by fibers counted, with the exception of AHERA (discussed below). Since the smallest fibers

are of questionable significance in producing toxicity, this enhanced sensitivity is not a great advantage except for purposes of research or where the concentrations are so low that an extremely sensitive technique is required to detect any fibers at all. A major difficulty encountered in this technique is the definition of a fiber, since at this high level of resolution asbestos fibers appear irregular, frayed, split, and often clustered. Electron microscopy is the usual technique for monitoring ambient air levels because of this sensitivity, but is too expensive and questionable for routine use in occupational hazard monitoring.

Recently, there has been widespread concern over the presence of asbestos in office buildings, public buildings, and especially schools. In response, the U.S. Congress passed the Asbestos Hazard Emergency Response Act (AHERA) in 1986, which required schools to inspect, monitor, and abate asbestos hazards. By now, almost all asbestos is out of schools, but occasionally an overlooked patch of asbestos is discovered. AHERA introduced a new set of rules for counting presumptive asbestos fibers and fiber complexes (using the terminology "structures") by electronic microscopy. The AHERA counting protocols are still widely used and AHERA accreditation has become the de facto qualification for technicians and experts dealing with asbestos exposure assessment in the community.

Asbestos in buildings may become airborne from friable insulating material by: (1) erosion of the insulation matrix, resulting in the release of embedded fibers into the air, (2) physical disturbance of the material, such as breaking it loose, and (3) stirring up deposited asbestos in carpets and on surfaces by activity, foot traffic, or cleaning. The first two are minimal in their effect. Erosion of the matrix is a slow process and it would take years for insulation material to deteriorate to such a degree that significant numbers of fibers would be released. Physical disturbance is primarily a short-term effect, releasing small clouds of fibers that are rapidly dispersed; the principal long-term effect of this release would be to hasten the deterioration of the insulation. As a practical matter, community risk is limited to free asbestos in the air and materials that contain asbestos in a softer matrix, such as sprayed asbestos-containing insulation coatings.

Harder materials, such as cement, pose no risk unless drilled or cut; and even then, the risk to occupants in the building would usually not be significant.

The most important community exposure to asbestos would be airborne exposure from secondary dispersion, when accumulated particles are stirred up by activity or air currents in a building with recent disturbance of asbestos. Routine housekeeping, such as vacuuming floors with HEPA filters in the equipment, should keep this to a minimum.

The recommendations of most authoritative bodies that have examined the issue are consistent. A selective program of building inspections to identify deteriorating asbestos-containing insulation should be followed by enclosing, sealing, or removing the asbestos once it has been shown to be friable. This is not only the most reasonable but also the most prudent course of action because it reduces the risk of inadvertently causing high dust levels by poor rip-out procedures. Visual inspection should be sufficient to identify hazardous conditions before a substantial amount of asbestos is released, and long before airborne levels become elevated, if carried out conscientiously and on a periodic basis.

In cities without local sites of asbestos release, typical urban levels have been on the order of 0.6 to 1.8 ng/m<sup>3</sup>, depending on the proximity to buildings under construction, and are declining as sources disappear. Fiber counts in this range must be made by electron microscopy because phase contrast microscopy is insensitive at such low levels; however, conversion to fiber counts would yield approximately 0.0002 to 0.0006 f/cc if counted optically. Despite such low levels in ambient air, persons living in an urban environment do accumulate a detectable body burden through the activities of daily life. Autopsies conducted even on urban residents without any asbestos exposure on the job demonstrate the presence of asbestos bodies and occasionally small, uncoated asbestos fibers, which are much harder to see, in the lung. Even rural residents typically have some fibers. However, the counts are very low for non-occupationally exposed persons, usually ranging below 100/gram in the lung tissue.

## CARBON MONOXIDE

Carbon monoxide (CO) is a colorless, odorless, and nonirritating gas that is the most commonly fatal toxic hazard in the community. Its effects are well known to most physicians. However, there are subtleties to CO poisoning that have only recently come to be appreciated.

In the community, CO is generated wherever there is combustion with a rich fuel-to-air ratio and oxygen deprivation, such as a smoldering fire or a low flame. It is heavier than air and is particularly dangerous in confined spaces, where it may accumulate to high concentrations. In the ambient environment, CO is generated by gasoline-powered, internal-combustion engines (much less by diesel), and is an important constituent of ambient air pollution. In the built environment, it is encountered where there is an open flame (such as a fireplace with the damper partly closed) or a poorly maintained furnace. Cooking on an inefficient open flame (such as a barbecue or hibachi, neither of which should ever be used indoors) can also generate CO, as can propane stoves, automotive exhaust (typically in a closed garage or communicating garage through which it can enter a living space), or an air compressor or similar fueled device used indoors.

In occupational settings, CO may be encountered in the context of exposure to motor vehicle emissions, incomplete combustion, or absorption (usually by inhalation) of methylene chloride (dichloromethane,  $\text{CH}_2\text{Cl}_2$ ), which is metabolized to CO in the liver. (See the section on solvents in this chapter.) CO is a particular hazard in fires, as a product of incomplete combustion, and therefore represents hazard to firefighters and fire victims. Police directing traffic, automotive or truck mechanics, and other occupations involving fire or motor vehicles are at greatest risk.

Furniture strippers, theater set construction workers, and others whose jobs involve stripping paint may be exposed to high concentrations of dichloromethane (methylene chloride ( $\text{CH}_2\text{Cl}_2$ )), which is metabolized by dehalogenation to CO. (Dichloromethane is also discussed in the section on solvents.)

CO, once inhaled, passes efficiently across the alveolar-capillary barrier and enters the bloodstream directly, where it is distributed throughout the body. About 90 percent remains bound to hemoglobin, and about 10 percent is extravascular. CO is not metabolized and is eliminated only by dissociation from hemoglobin, when it is displaced by oxygen, and by release, exiting the body in expired air. It follows first-order toxicokinetics, with a  $t_{1/2}$  of 13 hours.

CO has an affinity for hemoglobin 245 times greater than that of oxygen. CO therefore binds to hemoglobin quickly and almost completely, and competitively prevents oxygen from occupying the binding site. This reduces the capacity of blood to deliver oxygen to tissues. Because of its stearic effect on the hemoglobin monomer, which shifts the oxygen dissociation curve to the left, CO also interferes with the release of oxygen at the level of the tissue. The net effect is progressively less oxygenation of tissues with increasing accumulation of CO in the form of carboxyhemoglobin.

Anoxia has long been thought to be its only important mechanism of action. However, CO has been identified as a central nervous system neuromodulator, one of a class of neurotransmitters released in the central nervous system that diffuse rapidly to surrounding cells and have an inhibitory effect on activity. Exogenous CO may interfere with central function. Either through anoxia or directly as carboxyhemoglobin, CO exerts a vasodilatory effect on the cerebral circulation and is involved in active hyperemia, autoregulation, hypoxic dilation, and counteracting vasoconstriction. CO also binds to cytochrome oxidase and to myoglobin and may indirectly lead to lipid peroxidation and the production of intracellular reactive nitrogen species. Following CO exposure at toxic levels, withdrawal from CO may induce reperfusion injury and hypotension (due to vasodilation), compounding the toxic injury.

A consequence of the high affinity of CO for hemoglobin is that, over time, the level of carboxyhemoglobin rises with continued exposure as it is accumulated at the expense of oxygenated hemoglobin. Duration of exposure is as important as the level of exposure to CO because carboxyhemoglobin accumulates over time. High blood

concentrations may occur as easily with prolonged exposure to low levels as to transient exposure to moderately high levels. Ventilation patterns also play a role in the exposure; higher minute ventilation results in increased accumulation. Significant elimination of CO occurs only when the atmospheric levels are low. Inhaled CO follows a strict mass effect: the amount of CO in the body is determined, when the atmospheric concentration is elevated, by the product of concentration in the air, ventilatory volume (not rate) over time, and duration of exposure. Nothing else affects the determination.

Cumulative exposure to CO is measured by carboxyhemoglobin (COHb) in the blood, expressed as a percentage of hemoglobin, or, in conscious patients, by CO in expired air. Symptoms correlate roughly but not closely with COHb level. Measurements of arterial oxygen tension ( $P_aO_2$ ) may appear to be within the normal range, because the remaining oxygen that is associated with hemoglobin dissociates at a normal partial pressure. The usual signs of hypoxemia, such as cyanosis of the lips and nails, are not present because the cherry red color of carboxyhemoglobin provides almost normal coloration. At very high levels of exposure, vasodilation produces a mottled effect on the skin called "livor."

The symptoms associated with carbon monoxide toxicity include the full range of effects of hypoxia and reflect the blood concentration of carboxyhemoglobin. Table 10.4 presents the clinical exposure-response relationship for CO. Initially the effects of CO exposure are associated with headache, loss of concentration, and visual changes, including loss of acuity and tunnel vision (especially if there is hyper-ventilation). Hallucinations have been reported. Nausea and peripheral vasodilation may occur, resulting in syncope or lightheadedness. At higher carboxyhemoglobin levels, rising from 30 to 50 percent, loss of consciousness, lactic acidosis, convulsions, and death may result. Because CO interferes with oxygen delivery, it reduces oxygen supply to the myocardium as well as the brain. The result may be cardiac ischemia in persons with preexisting coronary artery disease; these changes may occur due to CO alone above 30 percent

**Table 10.4.** Clinical Exposure–Response Relationship for Carbon Monoxide

Carboxyhemoglobin (COHb) Level (%)	Characteristic Features (Toxicome)	Notes
1–5	Normal (nonsmokers approximately 1 percent, smokers up to 9–10 percent for two-pack/day habit)	COHb level reflects recent smoking
5–10	Cognitive impairment on neurobehavioral tests; usually asymptomatic, may feel headache	Smokers usually are asymptomatic at this level. Elevated risk of myocardial infarction, increasing with exposure.
10–20	Flu-like syndrome; headache, dizziness, nausea	Blood gases may show acidosis.
20–40	Tunnel vision, tachycardia, hypertension	Treat with supplemental oxygen
≥40	Obtundation, symptoms of anoxia. Cherry-red livor (mottled appearance of skin due to vasodilation), variable.	Treat with hyperbaric oxygen. Risk of post-recovery impairment
≥60	Skin bullae, coma, seizure, hypotension, death.	Multiple organ failure

carboxyhemoglobin. Induction of angina and increased frequency and complexity of arrhythmias have been demonstrated at levels as low as 6 percent in subjects with coronary artery disease. Thus, one of the most serious health effects of even low-level exposure to CO is the risk of angina, ventricular arrhythmia, and possibly myocardial infarction

in workers who may have silent or diagnosed coronary artery disease. Sudden exposure to very high levels may be fatal in minutes with no warning, due to chemical asphyxiation.

Severe and prolonged CO toxicity may result in focal hypoxic brain damage, which may later manifest itself as Parkinsonism (because it disproportionately affects the basal ganglia) and generalized cerebral edema. There is a high risk of a delayed neurological sequelae following recovery from acute CO toxicity severe enough to be associated with loss of consciousness, and more likely to occur in older victims. One syndrome occurs after a symptom-free recovery period of 20–30 days and resembles a stroke. It is associated with lesions of white matter in watershed regions, with poor blood supply, and may reflect hypotension during recovery. In children, survival from a symptomatic episode of CO toxicity may be associated with cognitive impairment. In adults or children, Parkinson's disease and other movement disorders, cortical blindness, paralysis, stroke-like symptoms (ataxia, apraxia, agnosia), and peripheral neuropathy may occur.

In occupational settings, these effects seldom act without modification by personal susceptibility factors, of which the most important is smoking history. Non-smoking adults normally have carboxyhemoglobin levels at about 1 percent and develop symptoms when their levels rise, variably, above approximately 5 percent. (Patients with hemolytic anemia have higher baseline levels.) Heavy smokers may not feel symptoms and may perform normally with levels of 5–10 percent, at which non-smokers would demonstrate cognitive impairment on neurobehavioral testing. Tolerance to higher carboxyhemoglobin levels renders smokers less susceptible to the effects of CO, at least at lower concentrations.

Management of CO toxicity is based on competitive displacement of CO from hemoglobin by oxygen, and re-oxygenation of tissue, especially the brain and myocardium tissue. In order to overcome the greater affinity of CO for hemoglobin, oxygen should be given at 100 percent, which reduces the  $t_{1/2}$  to about one hour, or at hyperbaric pressures if the patient is gravely ill and a chamber is available. In

extreme cases, exchange transfusions have been tried. Acidosis should be corrected. In fire victims and firefighters, the possibility of concomitant cyanide toxicity should also be considered.

## CYANIDE

Cyanide (CN) is a colorless gas that is lighter than air and is perceived by those with the genetic capability to have an almond-like odor. It exists in several forms, including the gases hydrogen cyanide and cyanogen chloride; as soluble potassium and sodium cyanide salts; as insoluble mercury, copper, gold, and silver cyanide salts; and as halogenated cyanides. It also exists as moieties in cyanogenic glycosides, which are natural products of constituents in plants such as cassava, lima beans, pitted fruits, and bamboo shoots. Hydrogen cyanide can be produced by the contact of cyanide salts, such as sodium or potassium cyanide, with acids or acid salts; by the hydrolysis of other cyanide gases such as cyanogen chloride; and by the hydrolysis of cyanogenic glycosides in foods and in the seeds of some trees (apricot, bitter almond, cherry, peach). Toxicity from cyanide may therefore occur in the context of exposure to cyanogen ( $\text{C}_2\text{N}_2$ ), cyanogen chloride (CNCl), hydrogen cyanide (HCN), or acrylonitrile ( $\text{C}_2\text{H}_3\text{CN}$ ), in roughly that order of toxicity. Cyanide is also formed by the metabolism of acetonitrile and of propionitrile.

Cyanide poisoning has become a subject of increasing interest in recent years because of the recognition of the important role cyanide may play in smoke inhalation-associated deaths, and the risk that cyanide might be used as a chemical weapon by terrorists. Another reason for the increasing interest in cyanide poisoning is the availability of a novel antidote, hydroxocobalamin, which can be administered in the pre-hospital as well as hospital settings.

Gaseous and solid forms of CN are both used extensively in industry, in metallurgy; electroplating; metal cleaning; recovery of gold and silver from mineral ores; recovery of silver from photographic materials; the production of plastics, pigments, and dyes; as a pesticide (fumigant and rodenticide); and in gold jewelry making. Other sources of CN

include emissions from the production of iron and steel, coal burning, vehicle exhaust, and cigarette smoke. Moreover, hydrogen cyanide is released as a product during the combustion of plastics (particularly nitriles) and natural polymers, including silk, wool, and cotton. Sodium nitroprusside, used in critical care to lower blood pressure, generates cyanide in the body.

Cyanogenic plant constituents are more likely to cause chronic rather than acute toxicity when ingested over time, and are typically not responsible for acute cyanide toxicity. The bio-accumulation of cyanogenic glycosides in the palm nut by bats, which are eaten in large numbers by humans, has been proposed as the mechanism of the historically extremely high incidence of tropical ataxic neuropathy in Guam, a disorder resembling amyotrophic lateral sclerosis that is only found on the island. A quack medical remedy of the past, Laetrile, was a cyanogenic compound derived from peach pits. Cyanogens are removed from many foods (cassava, flaxseed) or have been selected out of commercial strains (as in lima beans).

Occupations with potential exposure to cyanides include fire-fighters, chemical workers, custom jewelry makers, artists working in gold and silver, precious metal recycling workers, and workers exposed in agriculture. In the 1990s in Illinois, a CN-related death was prosecuted as murder in light of evidence that an employee of the firm Film Recovery Systems (the defendant in the case) died after inhaling CN while working in a plant owned by the firm. Other contributing evidence showed that plant workers were not informed that they were working with CN, that the plant was poorly ventilated, and that workers were not provided with protective clothing.

Depending on its form, cyanide may enter the body via inhalation, ingestion, or skin absorption. CN is not ionized, and it crosses membranes easily. It is distributed quickly throughout the body and has no storage depot. The body can detoxify small amounts of cyanide by several mechanisms, including metabolism in the liver by the enzyme rhodanese to the nontoxic metabolite thiocyanate, which is then metabolized or excreted in the urine. This pathway is easily

overwhelmed by exposure to any appreciable level of CN. CN binds to cystein and to red blood cells, but does not affect hemoglobin.

CN is thought to cause toxicity by inactivating mitochondrial cytochrome oxidase. CN binds to ferric iron ( $\text{Fe}^{3+}$ ) of cytochrome oxidase  $a_3$  to inhibit the oxidative function of this enzyme and thereby to block cells from using oxygen, which is the substrate of normal cellular respiration. This, in turn, causes cells to switch from aerobic (or oxygen-dependent) metabolism, which yields the cellular energy source adenosine triphosphate (ATP), to anaerobic (or oxygen-independent) metabolism, which generates lactic acid instead of ATP. Cell death occurs because of anoxia. Because cells are unable to extract oxygen from arterial blood, CN toxicity is characterized by a high oxygen saturation of venous blood. The heart, brain, and liver are particularly vulnerable to CN poisoning because of their high oxygen requirement.

CN is also highly irritating to mucous membranes and causes eye and throat irritation.

Early symptoms of acute cyanide poisoning include neurologic manifestations such as giddiness, confusion, headache, and dizziness; nausea and vomiting; palpitations, hyperventilation, or shortness of breath; and eye irritation. Later symptoms of acute cyanide poisoning reflect neurologic, respiratory, and cardiovascular depression arising from inability to compensate for hypoxia. Seizures, coma, respiratory arrest, and cardiac arrest can occur within minutes after exposure to moderate to high concentrations of cyanide. Exposure to moderate to high concentrations of cyanide can cause loss of consciousness in seconds, and respiratory depression and cardiac arrest can follow within minutes.

No laboratory test, including measurement of blood cyanide concentrations, can confirm the presence of acute cyanide poisoning within the time period required to initiate intervention. CN poisoning is therefore diagnosed clinically in the first instance, on the basis of signs and symptoms: rapid loss of consciousness or development of coma and cardiovascular instability in circumstances consistent with the possibility of cyanide exposure, such as a fire. The presence of ele-

vated concentrations of plasma lactate, a toxic by-product of anaerobic metabolism, in addition to impaired consciousness and cardiovascular instability, increases confidence in the diagnosis. Variable signs include cherry red complexion and bright red retinal veins and arteries. Cyanosis is absent because of the elevated venous oxygen levels.

Management of acute CN poisoning fundamentally requires removal from exposure, supportive care, administration of 100 percent oxygen, correction of acidosis, and management of complications. Anticonvulsants are given for seizures. Continuous infusion of epinephrine may be required to prevent hypotension. Cyanide exposure presents a hazard to rescuers and treating healthcare professionals, in expired air, vomitus, and clothing. Decontamination may be necessary.

There are two specific antidote strategies used in the treatment of cyanide toxicity.

The Cyanide Antidote Package (also known as the Cyanide Antidote Kit [CAK], Lilly kit, Taylor kit, and Pasadena kit), consists of amyl nitrite (available as pearls), sodium nitrite, and sodium thiosulfate. The rationale behind this kit is to induce methemoglobin, which binds CN with greater affinity than cytochrome oxidase, in order to restore aerobic respiration. The sodium thiosulfate binds the cyanide ion to form thiocyanates, which are much less toxic than cyanide and are readily excreted by the kidney. However, the logic of this treatment has been questioned. CN binds to methemoglobin relatively slowly, and reduction of the total hemoglobin available may reduce the oxygen-carrying capacity of the blood, which may be especially critical if there is also CO toxicity, as may be common in fire situations. Sodium nitrite and amyl nitrite in the CAK may cause potent vasodilation, leading to hypotension and insufficient perfusion. Cyanide-poisoned victims have marked hemodynamic instability already, making this risky. The sodium thiosulfate component of the CAK is limited by its slow onset of action, which is why it cannot be used alone.

Hydroxocobalamin is a precursor of cyanocobalamin (vitamin B<sub>12</sub>) and is given in huge doses, in grams. The form of hydroxocobalamin used to treat vitamin B<sub>12</sub> deficiency is too dilute to be used as a cyanide antidote. Hydroxocobalamin detoxifies CN by binding with

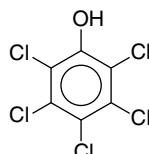
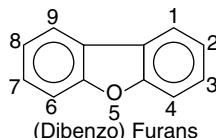
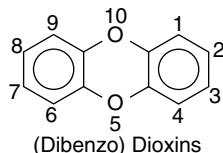
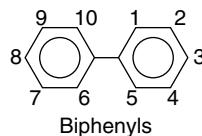
it to form  $B_{12}$ , which is then excreted in urine. There is significant intra-individual variability in toxicokinetics but the  $t_{1/2}$  in circulation is about 26 hours. It is well tolerated but has the unusual side effect of turning mucous membranes, skin, and urine red, and it may interfere transiently with specific colorimetric clinical laboratory values. Use of hydroxocobalamin has also been rarely associated with allergic response and anaphylactic reaction.

## DIOXINS AND FURANS

These terms refer to two classes of dibenzo compounds, consisting of two linked rings. The dioxins are a class of chemical compounds characterized by two six-carbon aromatic ring structures joined by two oxygens. The furans are an analogous class consisting of two six-carbon rings joined by one oxygen bridge and one carbon-carbon bond. The specific compounds in the class are determined by the number and placement of chlorines in the rings (see Table 10.5 and Figure 10.1). Analogous isomers of dioxin and the furan are comparable in toxicity. The toxicity of isomer types varies widely. In each case, the most toxic isomer is the 2,4,7,8-tetrachloro form, either as the (dibenzo) dioxin (TCDD) or (dibenzo) furan (TCDF). In colloquial language, when people refer to “dioxin” they usually mean TCDD.

**Table 10.5.** Possible Dioxins and Furans

Prefix	Number of Chlorines	Possible Isomers
Mono	1	2
Di	2	10
Tri	3	14
Tetra	4	22
Penta	5	14
Hexa	6	10
Hepta	7	2
Octa	8	1



**Figure 10.1.** Polychlorinated ring structures: polychlorinated biphenyls, dioxins, and furans, and comparison with pentachlorophenol.

Dioxins and furans are generated under conditions of low-temperature combustion by chemical reactions under oxidative conditions that result in condensation of chlorinated ring compounds, such as the PCBs, hexachlorophene, and pentachlorophene. They have also been contaminants of manufactured phenoxyacetic herbicides and hexachlorophene, and famously of the mixed herbicide Agent Orange used for deforestation in the Vietnam War, as well as other “Agents” consisting of different mixtures. Traces of the dioxins and furans are found wherever organic material and a chlorine donor are present in a form in which chemical reaction can take place: fireplaces (produced in high concentration when pentachlorophenol-treated wood is burned), incinerators, effluent from kraft-process pulp and paper mills, and cigarette smoke.

In recent years the technology for measuring the dioxins, in particular, has greatly improved. The distribution of the dioxins and furans in the environment is now appreciated to be much wider than was previously supposed. The dioxins are very photosensitive and tend to degrade quickly in the presence of light. In the absence of light, however, both the dioxins and the furans are very stable and require incineration at very high temperatures or chlorine stripping by chemical reactions to be broken down.

The dioxins and furans are usually absorbed by the oral or transdermal route, less often by inhalation of contaminated particles. They are very lipophilic and are readily stored in adipose, liver, and nerve tissue with minimal turnover and mobilization. They are potent inducers of mixed function oxidase activity, although they do not themselves appear to be readily metabolized or to require activation for their toxic effects. Excretion is primarily biliary and renal, but the agents may appear in the breast milk of lactating women who have been exposed. This appears to have been an important elimination route in women in Vietnam.

The dioxins and furans are potent toxins in rodent and other animal studies. They produce, in microgram (per kilogram body weight) dosages, an exceptionally wide range of effects. Hepatotoxicity includes acute yellow atrophy of the liver, porphyria cutanea tarda (and its dermatologic sequelae), and hepatocellular injury. Hyperlipidemia results, including elevation in cholesterol and triglycerides. Motor and sensory neuropathies occur. Teratogenic effects, such as cleft palate, occur at lower exposure levels. Abnormalities in cellular immunity result. Carcinogenesis occurs across a wide spectrum of sites. In short, these compounds deserve their reputation as highly toxic agents in animal systems.

Surprisingly, their effects in humans appear to be much more modest. Several studies have now been conducted of occupational and accidental community exposures to the dioxins and furans, the most famous being the Seveso incident in Italy in 1976, the study of U.S. veterans of the Vietnam War exposed to Agent Orange, and the Yusho incident of furan-contaminated cooking oil in western Japan

in 1968. Aside from chloracne (described in Chapter 17), few effects, and little excess of cancer or birth defects have been confirmed. The dioxins and furans are a good illustration of the sometimes extreme differences among species with regard to susceptibility to toxic agents.

The difference in toxicity is explained by the affinity of the dioxins and furans for Ah, a receptor molecule that is involved in migration of steroid hormones into the cell nucleus. Polycyclic aromatic hydrocarbons and polycyclic organochlorine hydrocarbon compounds bind to Ah. The affinity of dioxins and furans for human Ah is much less than for rodent Ah.

The Yusho outbreak did demonstrate severe neurological disorders, particularly in children, following massive exposure to furans. It is therefore clear that with high exposure levels, dioxins and furans can cause human disease.

Dioxins and furans have been evaluated for toxicological similarity to TCDD and have been assigned toxic equivalency factors (TEFs, where TCDD = 1.0; a different dioxin might be 0.1 or 0.001, lower TEFs being more common).

Dioxins and furans are very bio-persistent but they are metabolized slowly in the environment. The environmental impact of these compounds is very great, as they are ubiquitous pollutants potently toxic to other animal species and are readily bio-accumulated in the food chain, leading to high concentrations in certain ecosystems, especially in northern latitudes.

## HYDROGEN SULFIDE

Hydrogen sulfide ( $H_2S$ ) is the second most common cause of fatal gas inhalation exposures in the workplace, mostly in the oil and gas industry, sewers and wastewater plants, swine confinement facilities, manure collection systems, pulp and paper operations (in the kraft and sulfite processes), asphalt roofing, pelt processing, and any confined spaces in which organic material has decayed.  $H_2S$  is a potential community hazard in areas with geothermal activity and hot springs,

such as Rotorua, New Zealand, where there is volcanic activity, and near high-sulfur ("sour") natural gas wells and facilities.

The mechanism of action of H<sub>2</sub>S is often attributed to inhibition of cytochrome-*c* oxidase, and therefore to a mechanism identical to that of cyanide, causing asphyxiation at the cellular level. There are other mechanisms, however, including irritant effect on the eyes and lung tissue and reflex pathways in the lung. Pulmonary edema and mucosal irritation, particularly of the eye, are nonspecific irritant effects of hydrogen sulfide, reflecting its chemical reactivity. Brain sulfide levels appear to mediate some of the effects, possibly the abrupt loss of consciousness. Hydrogen sulfide may also be a neuro-modulator, and therefore toxic exposure may involve direct functional effects as well as cell toxicity.

H<sub>2</sub>S is poorly soluble in water, and exposure occurs only by inhalation. H<sub>2</sub>S dissociates, and the sulfide ion binds to heme compounds and is widely circulated. Sulfide is then metabolized by oxidation to sulfate, disappearing into the abundant sulfate stores of the body. Any excess sulfate in the circulation is excreted by the kidney, but inhalation of H<sub>2</sub>S contributes little to sulfate loading.

H<sub>2</sub>S has a foul odor, best described as the smell of rotten eggs, with a low odor threshold (0.01 to 0.3 ppm). Ordinarily this would be a good warning property, but H<sub>2</sub>S has an effect on the olfactory mechanism that shuts down the ability to smell it at levels that approach serious toxicity (100 ppm). As exposure increases, some habituation may take place, as with any odor, but the odor becomes nauseatingly intense at 5 ppm. It then disappears because of paralysis of the olfactory nerves. The conventional wisdom in oilfields, then, is that a worker is safe as long as he (almost all are male) can smell it, but is in danger when he cannot. This is the reverse of other odiferous exposures.

H<sub>2</sub>S does not follow Haber's Law (see Chapter 2); concentration is much more important than duration of exposure. The exposure-response curve for lethality is extremely steep for H<sub>2</sub>S. Current models for risk assessment use a concentration-time constant for lethality of the general form  $C^n \times t$ , where  $n$  ranges up to 4.36. Exposure to

1000 ppm is quickly lethal. Occupational exposure levels based on time-weighted averages do not consider this.

Acute central neurotoxicity, pulmonary edema, and the mucosal effects, especially eye irritation, are also characteristic features of acute toxicity of H<sub>2</sub>S. H<sub>2</sub>S has potent neurotoxic properties. Exposure at high levels results in a characteristic reversible sudden loss of consciousness called "knockdown" in the oil and gas industry at concentrations above 500 ppm. Workers who are knocked down may recover abruptly and without apparent immediate harm, although this condition may be fatal and may result in anoxia and head trauma. Many knockdowns go unreported, particularly in the field, as workers have been known to get up and finish the job without telling anyone what happened.

Pulmonary edema is also a well-recognized acute effect of H<sub>2</sub>S toxicity, especially when exposure is prolonged. Older studies suggest that 20 percent of cases of H<sub>2</sub>S toxicity reaching the emergency department showed evidence of pulmonary edema. However, compared to other causes of toxic inhalation, the prognosis for H<sub>2</sub>S-induced pulmonary edema is relatively favorable. Experimental studies have shown that although H<sub>2</sub>S at high concentrations produces a marked alveolitis and profuse edema, it is only moderately cytotoxic for pulmonary cells, and does not seem to disrupt the basement membrane of the alveolar endothelium. Thus, the ultimate prognosis for recovery and remodeling may be good if the patient is supported through the acute episode.

"Gas eye," or keratoconjunctivitis, is a superficial inflammation of the cornea and conjunctiva that is often recurrent in workers in sour gas plants who are exposed for prolonged periods to concentrations around 50 ppm. A peculiar feature of this effect is that it can be associated with reversible chromatic distortion and visual changes, possibly from transient deformities of the corneal epithelium that produce a colored halo surrounding an object. Keratoconjunctivitis may be accompanied by blepharospasm, tearing, and photophobia.

There is little data on teratogenicity, mutagenicity, or carcinogenicity for H<sub>2</sub>S, largely because it is so toxic. In animal studies, H<sub>2</sub>S

appeared unlikely to cause reproductive toxicity after exposures that are plausible in occupational and workplace situations. Evidence for cancer in human populations is weak. Studies of residents of Rotorua, which is a largely Maori community and cultural center, have observed an excess of cancers of the sinus overall (which may be confounded by wood dust exposure in Rotorua's extensive native craft industry), and of lung cancer in Maori women (which was not completely explained by higher smoking rates, but was accompanied by a statistically significant deficit among Maori men).

Chronic central nervous system effects following repeated or prolonged knockdowns have not been adequately studied. The available evidence strongly suggests that knockdowns are sometimes, if not usually, associated with chronic neurological sequelae. Extrapyramidal symptoms resembling Parkinson's disease may appear during the recovery from an H<sub>2</sub>S-induced coma. Experimental studies also help to localize the effects of H<sub>2</sub>S in the brain and seem to suggest anoxia and cytochrome oxidase inhibition as the mechanism for these effects.

Long-term, low-level toxicity is controversial for H<sub>2</sub>S. The literature on chronic, low-level toxicity of H<sub>2</sub>S is contaminated by numerous studies of questionable value and must be evaluated with great care. Although there is sufficient evidence to conclude that a chronic toxicity syndrome exists as a sequela of knockdown, the evidence to date is weak for the conclusion that a chronic toxicity syndrome exists because of long-term, low-level exposure. Ecological studies of residents of Rotorua suggest some excess morbidity in both the central and peripheral nervous systems, but the pattern is not consistent. It is strongest for mononeuritis, which is not plausibly a toxic effect, rather than for polyneuropathy.

Laboratory studies are generally not helpful. Sulfide measurements can be made on brain tissue post-mortem, but this is currently a research tool, not a forensic test. Blood sulfide levels are not diagnostic of H<sub>2</sub>S toxicity, are subject to many limitations, appear to be very transient, and are not available on an emergency basis. Urinary thiosulfate levels show promise as a biological exposure indicator in

the monitoring of occupational exposure, but are profoundly influenced by diet. The partial pressure of oxygen in arterial blood may be normal in H<sub>2</sub>S toxicity, unless there is pulmonary edema or another reason for respiratory compromise. Metabolic acidosis and a high artero-venous oxygen difference may indicate anoxia at the cellular level and may correlate with severity of toxicity, but this has not been documented.

Treatment of transient exposures, and of knockdowns in which the worker has regained consciousness, is not specific. Patients should be observed for pulmonary edema overnight, as with any toxic inhalation (see Chapter 14), and metabolic acidosis should be treated. Unlike cyanide, H<sub>2</sub>S does not require decontamination and does not present a hazard during resuscitation or emergency treatment.

Combined treatment with hyperbaric oxygen and nitrite is currently the therapy of choice, but the evidence to support it is anecdotal, as is often the case with uncommon toxic exposures. Given the low morbidity of hyperbaric oxygen treatment in skilled hands, it is a reasonable intervention if facilities are available.

Nitrite treatment has been recommended on analogy to cyanide, but the rationale is weak because the complex between sulfide and methemoglobin does not last long enough to make much difference, and sulfide disappears quickly from the circulation. A potentially life-threatening problem with nitrite is that it may add to the anoxia that accompanies pulmonary edema, and may also induce hypotension, further compromising the patient. Treatment with oxygen and supportive care alone has also been recommended in order to avoid further complicating the toxic effects with iatrogenic anoxia and nitrite toxicity.

Many other antidotes to H<sub>2</sub>S intoxication have been proposed but few have shown efficacy. Sodium thiosulfate, in particular, has been proposed as a treatment for H<sub>2</sub>S intoxication on analogy to its role in cyanide toxicity, but its efficacy has not been demonstrated and the rationale is in question. Sulfide is oxidized and excreted sufficiently rapidly by the body to rule out clearance as the problem. Thiosulfate is a metabolite of sulfide and not a substrate for enzymatic

oxidation destroying the toxic moiety, which is the case for cyanide intoxication.

Because H<sub>2</sub>S is heavier than air, workers entering a depression or confined space in which the gas has collected, or who are trapped in a plume escaping from a leak or well blowout, are at highest risk. Personal protection requires a self-contained breathing apparatus (SCBA). As in most such incidents, casualties usually occur at least in pairs, as would-be rescuers rush to save their co-workers and, in their haste, neglect to protect themselves with SCBA. Intensive training and ready availability of personal protective equipment are required to prevent such situations.

## LEAD

Lead is a natural element (<sup>207</sup>Pb), the metal product of radioactive decay. It is ductile, has a low melting point, high density, and unusual chemical properties, and it resists acid and corrosion.

Lead was probably the first metal to be smelted by humans and was extensively used in ancient times. The history of lead poisoning effects is almost as long. Hippocrates (370 BCE) was the first to describe the classical symptoms of lead poisoning: colic, pallor, paralysis, constipation, and ocular disturbances associated with lead toxicity from occupational and environmental exposures. Lead salts were used as a sweetener in wine by the Romans. Considerable exposure took place in the Middle Ages during smelting operations and probably in the production of leaded glass. The development of firearms introduced a new exposure opportunity with the production of lead shot. With the coming of the Industrial Revolution, the use of lead increased dramatically. Poor hygienic conditions found in industries of the day caused many cases of acute lead poisoning, and the metal became a widespread environmental contaminant. The use of lead has declined in recent years but it remains the most heavily used nonferrous metal in industry.

Occupational exposure still occurs and is still a problem, particularly in smelting and metal crafting operations in which lead is used for

stamping purposes and in lead battery recycling plants. New uses for lead and lead compounds have created many modern problems. The major use of lead in North America and in much of the world today is in the manufacturing and recycling of electric storage batteries. Lead also is encountered in electronic manufacturing in soldering processes, and is a major hazard of outdoor bridge maintenance and ship breaking operations. Exposure may occur from using firearms, loading ammunition, and working with some crafts involving glass and ceramics.

Lead is used in industry in many other forms, such as the metal, inorganic compounds and salts, and organic compounds. Inorganic compounds of lead are used in storage battery production. Lead is used in some outdoor paints. Lead compounds have become widely used as stabilizers in the production of polyvinyl chloride plastics. Lead-glazed earthenware improperly fired may dissolve lead in foods that are acidic. New technologies, such as ferrite ceramic magnets, which are piezoelectric compounds, ensure that lead will continue to be used.

Environmental exposure to lead is currently a problem primarily of the built environment. There are few natural sources of lead in the environment, although one example is the city of Kabwe, Zambia, where natural surface geological formations rich in lead have combined with the legacy of mining, smelting, and leaded gasoline to create an unusually severe problem.

Absorption of lead compounds depends on the chemical form. Lead in all forms is readily solubilized by nitric acid, but otherwise solubility in aqueous solutions varies with the anion and pH. Finely divided metallic lead is relatively soluble for a metal and, because the particles are small with a high surface area to volume, are readily inhaled and absorbed over time through the lung if they are retained. However, bulk metallic lead, in the form of bullets or shot, is not readily absorbed, although significant lead can be mobilized over time if a bullet lodges near a joint or in a highly vascular tissue. Ingested lead particles are absorbed poorly and transit the gastrointestinal tract before they can contribute much to internal dose. Lead salts, on the other hand, are readily absorbed by ingestion and are the

source of most lead toxicity. Lead compounds vary widely in bioavailability:

- Lead carbonate ( $\text{PbCO}_3$ ) is known as “ceruse” and, together with hydrocerussite ( $\text{Pb}_3(\text{CO}_3)_2(\text{OH})_2$ ) and “white lead” ( $(\text{PbCO}_3)_2 \cdot \text{Pb}(\text{OH})_2$ ), was used in white paint until 1978 and in outdoor weather-resistant paints afterward. Lead carbonate compounds are highly soluble and therefore readily absorbed; unfortunately, lead carbonate is attractive to children because it is sweet-tasting.
- “Venetian ceruse” was a white facial cosmetic used for beauty in the sixteenth century in Europe, despite its known toxicity; Queen Elizabeth I of England is portrayed in portraits wearing it. It contained lead carbonate as cerussite and hydrocerussite.
- Lead acetate ( $\text{Pb}(\text{CH}_3\text{COO})_3 \cdot 3\text{H}_2\text{O}$ ) is so sweet that it is called “lead sugar.” Lead acetate is highly soluble. It was used during the Roman Empire to sweeten wine and foods, which is the origin of suspicion that lead poisoning was rampant in Roman times. It is still used as a mordant for dyes, pigment, and print for textiles. Lead acetate is also an intermediate in the manufacture of other lead compounds.
- Lead (II) chloride ( $\text{PbCl}_2$ ) is not a salt (most metal chlorides are) and is insoluble in water; it is made by reacting lead with hydrochloric acid, which would usually dissolve and easily solubilize metals. It was used in white paints and is still used in leaded glass, nanotrees (nanostructures with an arboreal or dendritic structure), chemical catalysts, the production of metallocenes (organometallic compounds with catalytic properties), and as a fire retardant.
- Lead (II) oxide ( $\text{PbO}$ ) may exist in either of two forms: “litharge” or “yellow lead oxide” (orthorhombic), which is insoluble in pure water and increasingly soluble with lower pH in aqueous solutions, or “massicot” (tetragonal), which is poorly soluble in water. Both are used in lead storage batteries, ferrite ceramics, leaded ceramic glazes, lead glass, and yellow lead pigments.

- Trilead tetroxide ( $Pb_3O_4$ , the lead being in both divalent and tetravalent states), which is called “minium” or “red lead oxide,” is insoluble in water but readily solubilizes with increasing pH acidic aqueous solutions, and therefore is readily absorbed in the stomach. It was used as a pigment in red primer paints, in batteries, and in leaded glass (from which the lead can be transferred to acidic solutions).
- Lead phosphate ( $Pb_3(PO_4)_2$ ) is used in ferroelastic materials and has a wide range of solubility that is very pH dependent.
- Lead sulfide ( $PbS$ ) is “galena,” the most common lead ore, and was the dark pigment ingredient in “kohl,” a cosmetic still extensively used in the Middle East and south Asia. It is a semiconductor with thermoelectric properties and is used in infrared detectors and nanostructures.
- Complexed lead in the form of scale on the interior surface of leaded pipe is variably soluble depending on the redox potential of the water passing through. A high residual chlorine level or orthophosphate stabilizes the surface. A drop in residual chlorine, which occurred after many U.S. water utilities switched from chlorination to chloramines for drinking water disinfection in 2001, resulted in rising tap lead levels in homes served by old lead service lines before the problem was identified and corrected by the addition of orthophosphate.
- “Organic lead” compounds (the term technically should include lead acetate but usually does not) are rapidly absorbed.
  - Tetraethyl lead, which was the additive in leaded gasoline, is rapidly absorbed across the skin and has been associated with catastrophic lead poisoning in industry and, as a consequence of its use in gasoline, widespread environmental contamination.
  - Lead stearate ( $Pb(C_{17}H_{35}COO)_2$ ) is a lubricant and stabilizer for polyvinyl chloride (PVC) that is insoluble in water but readily absorbed by ingestion and transcutaneously. Lead can leach out of PVC products made with lead stearate.

Once lead enters the body it is transported to the various organs by the bloodstream, carried mostly bound to hemoglobin A<sub>2</sub>. Lead has an affinity for bone, which acts as a sink for 94 percent of the body burden of lead. Bone lead serves as a reservoir in equilibrium with blood, so that there are two elimination phases of toxicokinetics, a short phase and a long phase. The  $t_{1/2}$  for the predominant phase is 28 to 36 days in the adult. Lead is bound to bone in one of two compartments: the bone matrix, which binds the “old” lead for prolonged or permanent sequestration, and the surface of hydroxyapatite crystals, which contains the more available so-called “new” lead. The latter causes the most harm to the patient, since this “new” lead is more exchangeable. Mobilization of this “new” lead may result in toxicity even though months or years may have transpired since the last exposure to lead.

Inorganic lead is not metabolized. Organic (alkyl) lead compounds are dealkylated in the liver and remobilized as inorganic lead. Children have a greatly reduced capacity to metabolize organic lead compounds.

Excretion of lead follows two pathways; urinary excretion is the major pathway for environmental lead and typical occupational exposure levels. Biliary excretion augments elimination when exposure to lead is high, but is only partially effective because of enterohepatic circulation.

As little as 15  $\mu\text{m}$  absorption daily is necessary to achieve positive lead balance. Prolonged exposure to relatively low ambient levels of lead may therefore result in an excessive cumulative burden.

Lead toxicity in children has some common features with the lead toxicity found in adults, such as encephalopathy at high blood lead levels, but effects in the two groups differ based on level of exposure and internal dose. The neurotoxicity in children is usually central, while in adults it is mostly peripheral, with encephalopathy in adults occurring at higher levels. Since clinical lead poisoning is systemic and may mimic various other diseases, a definitive diagnosis requires laboratory confirmation of the clinical impression by obtaining a blood lead level reading in both adults and children. A high index of suspicion is most important.

## Children

Lead toxicity among children is primarily an issue of oral exposure in the home. Lead carbonate, which is highly soluble and therefore bioavailable, was used for many decades to make white paint. In older houses that have not been remediated, lead paint may flake and chip. When lead paint was banned in the United States in the 1970s (long after it had been banned in Europe), the mean blood lead level (BLL) in children abruptly dropped. Exposure to lead in the ambient environment is largely a legacy of leaded gasoline. Tetraethyl lead was used to boost energy yield from gasoline. When it was banned in North America, blood lead levels in American children dropped yet again. Lead in old paint and in house dust remains the most important source, even today. Occasionally additional sources are introduced into the home, such as lead-containing toys or contaminated foods. Other, usually minor sources of lead come from the drinking water distribution system in the home (see Chapter 12) and food.

In the United States, lead poisoning is not confined to children of low-income families; lead can be found in high-value houses undergoing renovation as well as in old houses that have been neglected. Lead has a sweet taste that many children find attractive as they orally explore their environment or engage in pica (the ingestion of dirt, clay, and other non-nutrient substances). Lead paint at the 1 percent level is considered acutely dangerous. Many pre–World War II buildings have paint containing as much as 40 percent lead. Not only lead from lead paint, but also lead deposited in soil from leaded gasoline in the past contributes to lead in house dust in the home. Concepts of lead toxicity in childhood have been drastically revised in recent years as lead levels have declined, revealing associations and background levels that were not known before.

For children, the Centers for Disease Control and Prevention (CDC) has established a so-called level of concern of 10 µg/dl to define elevated BLL in children. This is now universally recognized to be too high on a toxicological basis, because neurocognitive and behavioral effects have been documented on a group basis at lower

levels and show no evidence of a threshold for toxicity. Although the differences are small (about 2.5 IQ points for every increase of 10 µg/dl in BLL), the net effect of such a shift in a population that is normally distributed may be substantial at the tails of the distribution. (The same shift in mean IQ may theoretically quadruple the proportion of children with IQ < 80, for example, from 4 percent to 16 percent; at the other extreme, it may theoretically result in essentially no children with IQ > 125). In interpreting these data in terms of trend, however, it is useful to recall that lead levels are falling, not rising, and that past generations had much higher levels.

Other outcomes, such as elevated hearing threshold and decreased linear growth, impulsive behavior, and reduced academic performance, also show a correlation with BLL below 10 µg/dl. Therefore, the public health aim might logically be that levels of blood lead in children be as low as achievable, not simply below an arbitrary standard. CDC has retained the level of concern, however, because there is no feasible intervention, other than eliminating all sources of exposure to lead, that can be recommended to achieve levels lower than this.

There are two meaningful ways to monitor BLL in children. One is to track the percentage of children with  $\text{BLL} > 10 \mu\text{g}/\text{dl}$ , the CDC level of concern (because it is the threshold for public health surveillance, reporting, and case investigation), and the other is to track the geometric mean level of the population. The first approach reflects the frequency of higher-level exposure, such as ingestion of lead paint chips, that results in significant toxicity. The second reflects lead intake from all sources. Both are dropping but the patterns are different. The frequency of children with BLL is declining more slowly than the geometric mean, which suggests that on a national basis lead paint continues to be a serious threat, but that lead from other sources is playing a decreasing role. Blood lead levels in American children (1–5 years) have been falling for many years, from a geometric mean of 15 µg/dl in the 1970s to 1.9 µg/dl in 2002 (CDC data).

Children are much more susceptible to lead toxicity than are adults, in part because the blood-brain barrier is not mature in children, and because of their smaller size and immature nervous system. They

are also more vulnerable to pick up lead-containing dust because of hand-to-mouth behavior and floor crawling. Nutritional deficiencies of lead, calcium, and vitamin D increase the intestinal absorption of lead. Inborn errors of heme metabolism and hereditary anemias predispose children to the effects of lead toxicity. Maternal lead exposure can significantly expose the fetus in utero.

Acute lead poisoning in children is rare in modern times. In children, the onset is usually insidious with complaints of abdominal pain, anorexia, vomiting, constipation, and diarrhea. Anemia, resulting from decreased hemoglobin synthesis, occurs at about 40 µg/dl. Lead encephalopathy is a devastating condition, observed in children with BLL around and above 60 µg/dl, that is characterized by an odd combination of irritability and lethargy, with cognitive impairment. It carries a 25 percent risk of death and a 40 percent risk of permanent neurological impairment. Encephalopathy is most common in children under the age of four and may show a progression from drowsiness, lack of coordination, cranial and peripheral nerve palsies, seizures, and finally coma, if lead exposure continues to result in an accumulation of body burden. A lead line on the gums may be present beneath the base of the teeth as a result of precipitation with sulfide in the mouth ("Burtonian lines," resulting from gingival precipitation of lead bound to sulfide generated by mouth bacteria), and lead deposition may be evident on x-ray at the epiphyses.

The essential intervention in protecting children is to prevent contact with lead paint and house dust containing lead, through dust control and remediation of housing. Almost all jurisdictions have lead poisoning prevention operations that investigate lead sources when cases of BLL above the level of concern are identified. However, few have proactive programs that identify and remediate houses that may have lead interior paint before children are allowed to live in them.

## **Adults**

Adult lead exposure usually results from occupational activities and from inhalation. The CDC guideline for the diagnosis of lead

toxicity in adults is  $>80 \mu\text{g}/\text{dl}$ . Recent research suggests that this number, too, is too high, and that levels well below this are associated with a risk of declining cognitive ability and memory in the elderly.

Adult lead poisoning symptoms are similar to those found in children except that they occur at higher levels, and central neurological changes are rare. The signs of lead toxicity may be more subtle in their appearance in adults. Symptoms involving the GI tract usually appear with blood lead (BPb) concentrations slightly above  $80 \mu\text{g}/100 \text{ ml}$ . These symptoms are usually manifested as loss of appetite, epigastric discomfort after eating, and either constipation or diarrhea. As BPb concentrations continue to increase, symptoms become more severe, culminating in recurring spasms (colic). Muscle aches are common but may not be noticed until after they are gone, with treatment or removal.

The classical symptoms of adult lead poisoning (catastrophic toxicity) are gastrointestinal symptoms of thirst, metallic taste, nausea, vomiting, and intestinal colic (which is related to neuropathy of the intestinal enervation). These symptoms may progress rapidly to muscle weakness, pain, and paresthesias. After this acute attack, a persistent disease in the chronic form of lead poisoning usually follows. Lead toxicity affecting renal function usually results from heavy and prolonged exposure. Central nervous system toxicity from either acute or chronic lead poisoning may result in extreme cases in encephalopathy. This effect is observed at levels around  $80 \mu\text{g}/\text{dl}$  in adults. In the acute form there is edematous swelling of the brain, with most of the extensive neural injury found in the cerebellar cortex. The prognosis for complete recovery of neurological function after acute lead encephalopathy is poor, as glial cells proliferate around blood vessels. In chronic encephalopathy, the disorder is axonal injury. Subtle changes in mental attitude, memory deficits, and decreased ability to concentrate present first, progressing to drowsiness, stupor, and coma, if the encephalopathy is allowed to continue untreated and exposure continues.

Peripheral neuropathies attributable to lead poisoning require prolonged exposure and may result in paralysis or palsy due to selective involvement of motor neurons, with little or no sensory abnormalities.

Lead neuropathy is typically characterized by extensor weakness with wrist drop. Muscle weakness also occurs. This degree of lead toxicity is rare today. Neuropathies this severe have a poor prognosis. Chronic low-level exposure to lead may also result in multiple peripheral neuropathies, motor or mixed sensory and motor abnormalities, subtle cortical dysfunction, and mental status changes.

With improved hygienic conditions, renal complications due to lead ingestion have virtually disappeared, but remain a risk among consumers of illicit whiskey distilled from radiators in which lead solder has been used. The two principal functional impairments of the kidneys due to lead poisoning are the “Fanconi triad” of decreased re-absorption of amino acids, glucose, phosphate, and citric acid; and “saturnine gout.” Nephropathy, as an interstitial nephritis, results in reduced clearing of uric acid with associated hyperuricemia. Saturnine gout is a crystalline arthritis due to uric acid that is caused by decreased tubular function from lead toxicity. The condition presents like regular gout.

An appropriate clinical evaluation should concentrate on identifying these disease states, which are well-documented associations of lead toxicity; but the screening system to be employed should be comprehensive in order to identify other associations. The primary concern is identifying early signs of lead-related disease in workers with an exposure history at, or not far below, the range generally considered to be at risk (e.g., blood lead between 40 and 80  $\mu\text{g}/\text{dl}$ ). The demonstration of an elevated blood lead level does not in itself mean that the worker is suffering from clinical lead poisoning.

On the other hand, recent research suggests that there may be more subtle effects among adults with evidence of a higher long-term body burden than was previously realized. These may include a greater risk of dementia later in life. Lead levels are also associated with a higher rate of infertility. IARC (see Chapter 2) has determined that lead is a probable human carcinogen.

Under the OSHA Lead Standard, “medical removal” (removing the worker from exposure) is required if the BLL is  $>60 \mu\text{g}/\text{dl}$  on a single sample or  $>50 \mu\text{g}/\text{dl}$  averaged over the last three samples

repeated every month. BLL of 40–60 µg/dl requires repeated BLL readings every two months. Workers can return to work when the BLL falls to <40 µg/dl, with confirmation two weeks later that it has remained low and with repeated BLL determinations every 6 weeks. This is not a very aggressive intervention, and it obviously presumes that lead toxicity in adults is reversible and carries no long-term risk, a questionable assumption.

Exposure to lead in the general workplace can be controlled by housekeeping (keeping dust levels down), ventilation, personal protective equipment, and prohibition of eating, drinking, and smoking at work (to prevent inadvertent hand-to-mouth intake).

### Chelation

Chelation is an approach to treatment that involves a substance that entraps and solubilizes atomic lead, making it easier to transport the lead to the kidney for excretion. In adults and children with elevated BLL, the question of chelation is raised frequently—and often, unfortunately, inappropriately. Chelation is never a substitute for control of the exposure; and chelation while exposure continues may place the exposed person at increased risk by keeping a greater fraction of the lead intake in the circulation. Chelation may lead to increased absorption of lead if exposure continues and to increased retention if exposure is overwhelming.

Chelation is not a treatment to be undertaken casually. It involves mobilizing stores of lead that were previously sequestered, and the benefit in doing this must outweigh the potential harm. It is also an inefficient treatment, reducing body burden by only as much as 2 percent and not significantly in the central nervous system, where it counts most. Chelation may deplete the body of calcium and trace elements if prolonged.

The Centers for Disease Control and Prevention has promulgated guidelines for chelation for the management of lead toxicity, and the American Academy of Pediatrics has joined in these recommendations for children. Central to these recommendations is the discouragement

of excessive and premature use of chelation. Chelation for lead toxicity is secondary to removal from exposure and abatement of the home environment for children. It is not considered for children at levels  $<45 \mu\text{g}/\text{dl}$ , may be considered above that, and is recommended as a medical emergency at  $\geq70 \mu\text{g}/\text{dl}$ . Chelation is not recommended for adults at levels  $<80 \mu\text{g}/\text{dl}$  and is recommended above that level only if the adult patient is symptomatic.

Chelation with older agents can also be dangerous, resulting in calcium depletion, hypercalcemia, and nephrotoxicity from the solubilized lead. Agents used in past have included CaNa<sub>2</sub>-EDTA (calcium disodium EDTA—*never* EDTA alone, as this has resulted in fatal hypercalcemia), BAL, and D-penicillamine, the latter having emerged as the agent of choice, largely because it can be given orally. Each agent has potentially severe side effects, and none is currently recommended.

The current chelating agent of choice is dimercaptosuccinic acid (DMSA), known as Succimer. It can be given orally, and has fewer side effects than other agents; some side effects are rash, nausea and vomiting, and lightheadedness.

DMSA has replaced D-penicillamine, which was previously the oral chelating agent of choice for lead; CaNa<sub>2</sub>-EDTA, which was used when other agents were ineffective; and British anti-Lewisite (BAL), which is still used for rescue in cases of lead encephalopathy.

Chelation should be considered in symptomatic lead poisoning only under the supervision of a knowledgeable specialist and with careful monitoring of renal function. It should be undertaken with close monitoring, especially of lead levels and renal function.

## Anemia

Lead causes a number of hematologic effects, most characteristically a normochromic, normocytic anemia with basophilic stippling of red cells. The effects of lead on the heme synthesis are the only ones for which a satisfactory dose-response relationship has been established. Lead reacts with amino groups, carboxyl groups, and phenoxy groups,

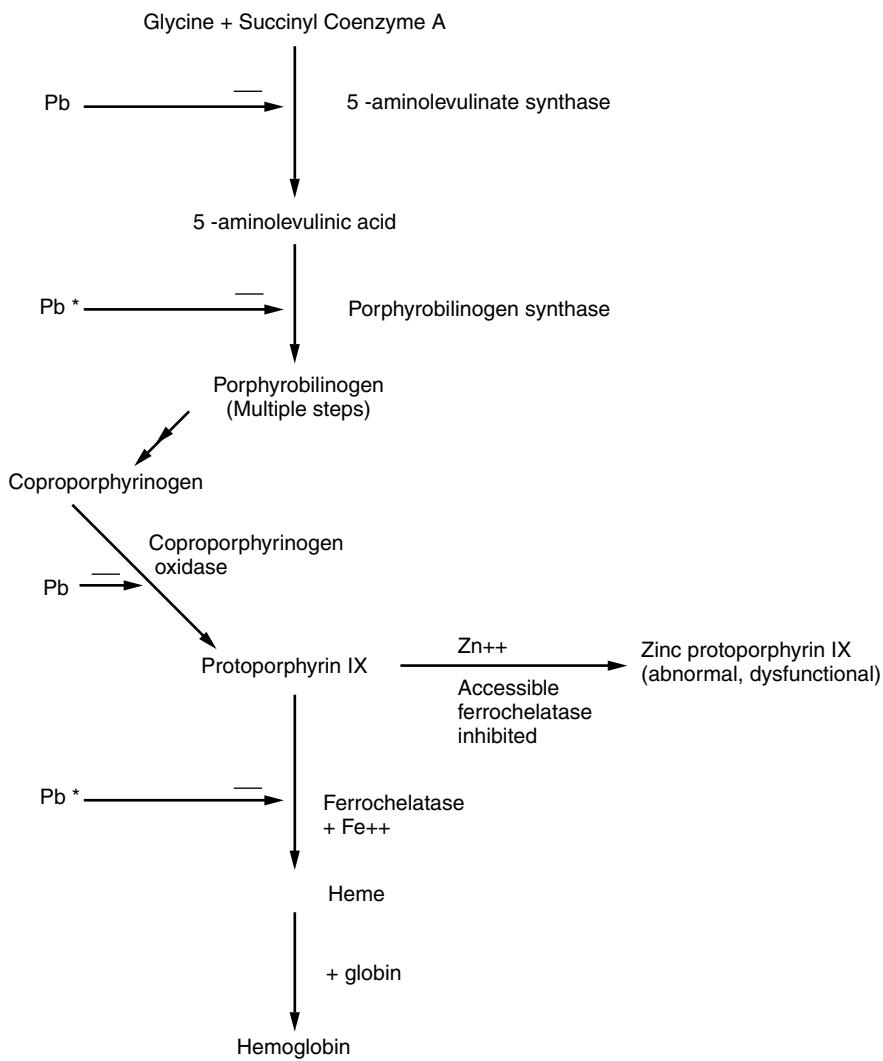
and alters the structures of macromolecules in addition to disrupting enzyme activity.

The effects of lead on the hematopoietic system (which causes anemia) are twofold: inhibition of hemoglobin synthesis, and shortened life span of circulating erythrocytes. This results in a compensatory increase in red cell production and the appearance of immature and hypochromic cells in the circulation.

The normal heme biosynthetic pathway is regulated by a negative feedback control (see Figure 10.2). As heme is produced, it depresses the activity of aminolevulinic acid (ALA) synthetase in an effort to overcome the block. One of the enzymes inhibited by lead is aminolevulinic acid (ALA) dehydratase. This results in increased activity of ALA synthetase. As the negative feedback system tries to compensate for this high ALA production, the amount of heme produced is further decreased. The next stage in the pathway that is clearly inhibited is coproporphyrinogen oxidation. The depression of coproporphyrinogen oxidase decreases the production of heme and results in an increase in coproporphyrin. Finally, lead inhibits the insertion of iron into heme by the action of ferrochelatase, reducing further the functioning heme available for incorporation into hemoglobin. The effect contributes to the development of sideroblastic anemia.

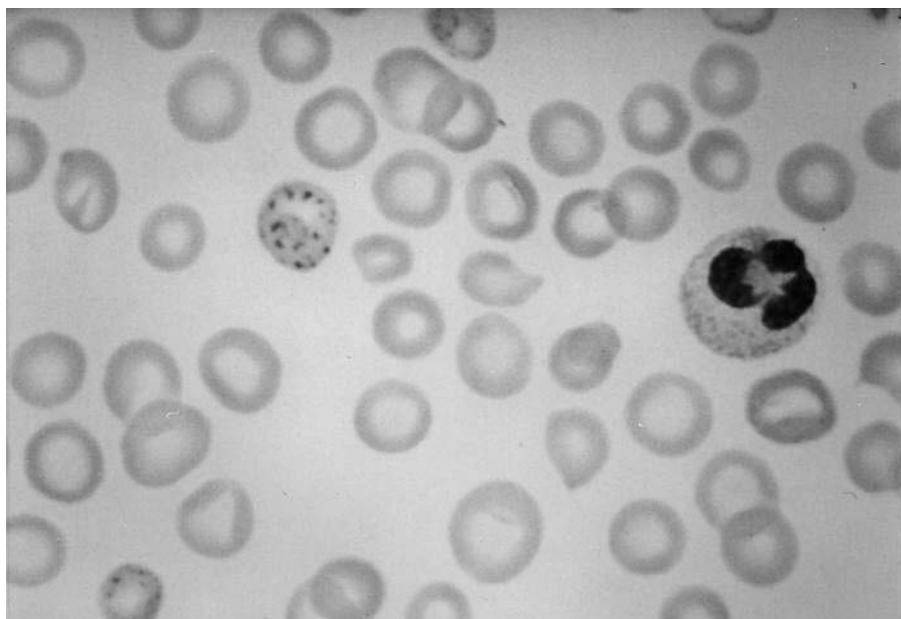
An important characteristic effect of lead toxicity on the biosynthetic heme pathway that has been useful in diagnosis is basophilic stippling of red blood cells (see Figure 10.3). Lead toxicity causes a deficiency of erythrocyte pyrimidine 5'-nucleotidase. This allows the formation and accumulation of pyrimidine nucleotides, which are usually virtually absent in normal erythrocytes. The accumulation of pyrimidine nucleotides has the effect of retarding the breakdown of RNA. This results in aggregation of undegraded ribosomes, giving rise to basophilic stippling, which is a characteristic sign of severe lead toxicity in the peripheral blood.

The second mechanism by which lead causes anemia is a shortening of the life span of erythrocytes. This shortening of the RBC's life span is probably due to direct toxic effect upon the cell membrane. Lead interferes with  $\text{Na}^+$ - and  $\text{K}^+$ -dependent ATPase activity, causing a



\*Hematologically significant effects

**Figure 10.2.** Hemoglobin synthesis and the effects of lead.



**Figure 10.3.** Basophilic stippling of erythrocytes as seen in a smear of peripheral blood.

major leakage of potassium. This loss causes dysfunction in the energy-requiring pump, since at the same time there is little change in the input of  $\text{Na}^+$  concentrations. Consequently, there is an intracellular rise in sodium and a potassium-mediated shrinkage of the cell. This leads to mechanical fragility of the red cell and consequent lysis. As more and more red cells are destroyed, together with inadequate hemoglobin synthesis, the patient becomes anemic. With this destruction of red cells there is greater demand placed on the bone marrow for the production of erythrocytes. This accelerated removal of red cells from the bone marrow stimulates erythropoiesis in the bone marrow. The erythrocytes produced are abnormal due to the hematologic changes and compensatory cellular proliferation in the absence of sufficient hemoglobin. Morphologically the red cell precursors are generally hyperplastic and present nuclear changes, mitotic abnormalities, megaloblasts, polyploid erythroblasts, and sideroblasts. There are more stippled cells found in bone marrow than in the peripheral blood.

The diagnosis of lead poisoning is often said to be difficult, but screening for lead toxicity is very easy. BLL has replaced all other diagnostic tests because it is readily available and sensitive. Earlier tests, such as zinc protoporphyrin and free erythrocytic protoporphyrin, are now obsolete. One major reason for this is that they are insensitive below 20 µg/dl.

## MERCURY

Mercury is an important occupational hazard, and an even greater environmental health issue. Mercury exists as a liquid elemental metal that sublimes to a vapor or in organic forms, which present somewhat different toxicity profiles.

The uses of mercury in health and dental care have presented a distinct challenge to hospitals, clinics, and physicians. Mercury has been used in thermometers, sphygmomanometers, and numerous other applications for instrumentation. Exposure to mercury from dental amalgams, in which gold is dissolved in mercury to form a pliable alloy for filling cavities, is thought by many to be a health hazard, and the search for evidence of this has constituted a major distraction in mercury toxicology. Historically, mercurial compounds, such as salvarsan, were antibiotics used to treat syphilis but had many side effects. Mercuric chloride, which is highly irritating, was used in the early days of sterile surgery as a disinfectant for hand washing; dermatitis caused by this disinfectant led to surgeons adopting rubber gloves.

Elemental mercury ( $Hg$ ) is a heavy metal that, like many others, binds to sulphhydryl groups of proteins, blocking enzyme active sites. Mercury is unique among metals in being volatile at room temperature, presenting a unique and insidious hazard. Mercury metal ( $Hg^0$ ) is the most common occupational and air pollution hazard. Inhalation is the primary route of occupational exposure to mercury metal and may occur in dental offices, health care institutions (typically because of broken thermometers), and schools (particularly science laboratories). In unusual circumstances, it can also occur in homes. Mercury is rapidly distributed to tissues and accumulates in liver and kidney,

which serve as depots. Mercury is not metabolized in mammalian tissue but can be methylated by fish and other aquatic species.

Organic mercury compounds are more rapidly absorbed than elemental mercury, especially by ingestion, and are more toxic than elemental mercury. They are the result of biotransformation in the environment and bioconcentration, accumulated by fish in mercury-contaminated rivers and estuaries, for example. Organic mercurial compounds have been used as fungicides, particularly as methylmercury on seeds. Organomercury compounds are metabolized to mercury in the body and therefore have the same toxicity as metallic mercury and additional neurotoxicity due to penetration of the central nervous system.

The major toxic manifestations of metal mercury are nephrotoxicity, primarily proteinuria and tubular necrosis, and neurotoxicity, which can be profound with high exposure. Oral and gingival inflammation accompany high-level toxicity. Symptoms include tremors, ataxia and gait disturbance, excessive and uncontrollable salivation, visual changes, anorexia and weight loss, labile affect and irritability (*erethism mercurialis*) associated with pathological shyness and avoidance of people, and, in children, acrodynia (erythema and pain in the fingertips, accompanied by desquamative dermatitis of the hands and feet). Deterioration of handwriting over time is often used to monitor progress of the disease, and a comparison of handwriting samples over months and years may reveal an earlier subclinical effect.

Mercury is often compared to lead in its pattern and potency for neurotoxicity. The two metals appear to be similarly neurotoxic, to impair cognitive development in children, and to be associated with behavioral changes. However the neurological effects of mercury tend to be more specific and focal, leading some to compare lead to a shotgun and mercury to a rifle in the specificity of their effects on the central nervous system.

As an environmental hazard, emissions from the burning of coal constitute the major loading of mercury in the environment, and became a public issue with enforcement of the Clean Air Act of 2008. The U.S. Court of Appeals for the District of Columbia overturned

an EPA rule permitting continued mercury emissions from power plants at then-current levels and delaying requirements for reductions. The issue has been revisited in the new administration.

Organomercury compounds (primarily methylmercury,  $\text{CH}_3\text{Hg}$ ) are now rarely encountered at overtly toxic levels, but have been important historically in waterborne mercury exposure, and remain important in ecotoxicity.

Methylmercury is bio-accumulated in aquatic species. At levels encountered in populations that depend on mercury-rich fish in their diet (most studied have been residents of the Faroe Islands and the Seychelles Islands), there is strong evidence for neurocognitive effects at high levels of consumption. There is also a more complicated effect on the heart in which cardiovascular risk from mercury is partly offset by omega-fatty acid intake, which is protective.

There is an unequivocal hazard to the fetus when fish with elevated levels are consumed during pregnancy. These fish species are usually top carnivores in the ocean ecosystem. The Food and Drug Administration, together with the EPA, has advised pregnant women, women who might become pregnant, and nursing mothers to follow these guidelines (paraphrased from the original joint announcement in 2004). The same guidelines should apply to the diet of young children, who should be served small portions:

- Do not eat swordfish, tilefish, king mackerel, or shark, because of elevated levels in these species.
- Eat up to but no more than 12 ounces per week during pregnancy, which corresponds to two “typical” servings, of fish and shellfish that are low in mercury: shrimp, canned light tuna, salmon, pollock, and catfish.
- Eat no more than 6 ounces of albacore tuna per week (because levels are higher than in other tuna varieties).
- Obtain advice locally on the mercury content of fish caught in local waters, for example in recreational fishing; but do not eat more than 6 ounces, and do not eat any other fish in the same week.

These are complicated guidelines. It is not clear how well they are followed, but the issue has received a great deal of media attention, and public awareness is high.

The fear of organomercury toxicity to the fetus is understandable. Congenital mercury toxicity may result in severe mental retardation and motor abnormalities, including disturbances in swallowing. Congenital mercury poisoning was poignantly illustrated by the severe outbreak in Minamata, Japan, in the 1950s, which was documented in *Minimata*, a famous photo-essay by W.E. Smith (the photographer) and A.M. Smith in 1975.

In dental offices, mercury is used to make amalgam for filling cavities. Some cases of overt toxicity in heavily contaminated dental offices have been reported, particularly following ineffective cleanup of spills. Most dental personnel are now aware of the hazard and often request blood mercury determinations as screening tests. Also, there is presently a debate over whether the minute amounts of mercury that volatilize from dental amalgams in patients may lead to toxicity. This has now been extensively studied, and the best evidence suggests that toxicity is unlikely due to dental amalgams alone.

In the home, mercury exposure has been associated with spillage of small quantities kept in the home for hobbies or as novelties, and as a by-product of poorly advised efforts at home smelting and gold refining. A severe case of home exposure, known but not reported in the medical literature, involved a quantity inadvertently spilled into a space heater; the heater became a source of exposure over many months. Washington, D.C., has had at least two recent episodes in which irresponsible students spread mercury in schools, requiring expensive decontamination. Levels of mercury in blood or urine do not correlate with symptoms, probably because the toxicity is intracellular, and much circulating mercury is bound with protein. However, cases of significant mercury toxicity will at least show elevated levels. Relatively low levels of exposure with urine or blood levels just outside the range of a reference population (varies by laboratory, 24-hour urine yields of 0–50 nanomoles per liter being typical) have not been associated with demonstrable abnormalities.

Hair analysis should be considered a research tool, as it is not reliable for clinical diagnosis; the preparation and analysis of hair requires careful preparation and is subject to inaccurate results from contamination.

Treatment of serious mercury intoxication should be left to the few physicians with experience in clinical toxicology. Chelation, with DMSA (Succimer) or, with a much higher risk of side effects, dimercaprol, penicillamine, or their analogues, is generally less successful in mercury than in lead toxicity, presumably because the mercury is more likely to be bound. The indications for chelation in cases of mercury toxicity are not as well defined as for lead. Chelation should be undertaken only in the presence of a certain diagnosis and under the management of an experienced clinical toxicologist.

## **MYCOTOXINS AND MOLD**

Mycotoxins are chemicals synthesized in mold, the end metabolic products of often unique pathways that presumably evolved for competitive advantage against other species of fungi and bacteria. They are included in this chapter not because they pose the threat of serious chemical exposures at the levels encountered in daily life or are frequent causes of disease, but because there has been extensive discussion of their role in human disease recently, and confusion over outcomes that result from allergy and that are ascribed to toxicity. Mold and mycotoxins have become major issues in litigation in recent years, and therefore the basic issues should be understood by OEM physicians. Additionally, the presence of trace mycotoxins in food is an ongoing concern in food safety, mostly from a regulatory point of view, because of cancer risk.

A review of the toxicology of individual mycotoxins is beyond the scope of this book and would ultimately be unrewarding. It is a highly specialized area of research of primary concern to food safety professionals, particularly those connected to the Codex Alimentarium, the UN-sponsored international monitoring and regulatory advisory body.

Molds are fungi and may exist in vegetative forms or as spores. In nature, molds have an ecological niche in soil, and they break down

dead or decaying vegetation, recycling organic matter such as cellulose in wood. Mold species vary in their requirement for water, the extent to which they survive drying, and their substrates. Mold species generally thrive on cellulose and in locations where there is persistent water.

The primary health effect related to exposure to mold is allergy. Mold is one of the four major allergen groups that predominate in patients with allergy and especially asthma. The other major allergen groups are dust mites, cockroach antigen, and animal proteins (especially cat). Mold species produce antigens in abundance, many of which cross-react. The usual trays for prick testing or the panels in RAST testing do not test for more than a handful of mold species and often miss mold allergy. Therefore, a negative allergy test does not necessarily mean the absence of mold allergy. Likewise, the standard panel for mold-related hypersensitivity pneumonitis features tests for sensitivity to only five antigens among the many that may be associated with the disease. Clinical judgment is therefore required to interpret a history of mold exposure and allergic response.

A secondary effect of exposure to mold spores may be mucosal irritation. Mold spores, and other bio-aerosols, carry several irritating chemical compounds, including  $\beta$ -glucans.

“Mold cases” can be evaluated by assessing the environment and by assessing the patient. The environment may be assessed (usually by consulting engineers and hygienists) by visualization, checking for moldy odor (“geosmia”), checking for moisture, and by bulk and wipe sampling, to determine the location of mold in the structure, and air sampling for airborne mold. Usually, the assessment identifies the presence of mold and bacteria, particulate matter in general, and quantifies the level of viable mold and identifies the species. It also determines whether there is “amplification” (mold growth indoors) by comparing levels inside and outdoors. *Cladosporium* is a very common mold genus typically found at higher levels in air than other molds. *Alternaria alternata* is very common. *Aspergillus* and *Penicillium* are common genera in air and on surfaces, and may indicate persistent moisture. *Stachybotrys chartarum* is a sticky, black mold not usually found in air—and, if it is, indicates a severe problem of persistent

moisture in a location that communicates with the air in the space being sampled. *Cryptococcus* species may indicate contamination with bird droppings.

Physicians get involved when residents seek medical evaluation for suspected mold-related disease. Molds cause illness by deep or superficial infection (mycosis), and by immunological response. Immune responses to mold may take the form of familiar allergies, especially asthma and rhinitis, or (in the context of inhalation of large quantities of persistent antigen such as spores) hypersensitivity pneumonitis.

In recent years, some investigators have proposed that human disease can also be caused by mycotoxins. Molds produce these and other chemicals, in part as an ecological defense. Some, such as penicillin, give mold a competitive advantage against bacteria. There are only two recognized human diseases attributed to mycotoxins, however. Human mycotoxicosis is a severe toxicity syndrome documented during famines in Eastern Europe when people consumed grossly contaminated grain and animals fed on them; symptoms are protean and sometimes lethal. Organic dust disease syndrome is an uncommon febrile illness that is an occupational disease of workers who inhale moldy grain dust.

The evidence that inhaled mycotoxins can produce human disease at levels observed in indoor air is weak. There are persuasive studies in the literature that describe an increased frequency of respiratory and allergic symptoms among residents of homes that are damp, but these have not shown an association with mycotoxin exposure. Concern over mycotoxins increased when the CDC conducted a study of an outbreak of pulmonary hemorrhage among children in Cleveland and reported an association with *Stachybotrys chartarum* in the home. On further investigation, the association did not stand up under scrutiny, and the CDC has now withdrawn its conclusion. Even so, allegations of “toxic mold” and “killer mold” continue. Public health agencies, including the California Department of Health Services, have found it necessary to release advisories correcting the public impression that an epidemic of mold or mycotoxin-related disease is in progress.

Some clinicians believe that mycotoxins produce a syndrome resulting in chronic illness, including persistent neurological impairment. There is no scientific evidence for this. Clinical tests for mold exposure are based on serology: antibody to specific molds can be measured using a special—but expensive—test available from specialty laboratories. These tests are not recommended because they serve little or no useful purpose. Molds are ubiquitous, and exposure occurs often. Titers typically rise following exposure and fall over weeks thereafter; this does not indicate an association with clinical disease, and there is extensive cross-reactivity among mold antigens. Tests for antibody against mycotoxins are even more expensive and cannot be interpreted. There is a panel of antigens intended to identify the likely cause of hypersensitivity pneumonitis, but it is useful only in the investigation of that disorder, which can be caused by many antigens not included in the panel.

In recent years, changing construction methods, water intrusion resulting in wet drywall, and increased awareness on the part of homeowners and tenants have led to concern over mold contamination in homes and other buildings. This has resulted in litigation, mostly for negligent construction and water intrusion. Some plaintiffs have won very large awards. Many of these awards are predicated on risks to health.

Unfortunately, the public preoccupation with mold-related disease has occasionally led to neglect of proper medical evaluation and inappropriate treatment. Mold and mycotoxins are not plausible explanations for unexplained or nonspecific symptoms or diseases. Allergies are by far the most common disorders caused by molds, and should be investigated in the normal manner. Physicians should be careful to not be misled by misplaced concern over “toxic mold.”

## NANOMATERIALS

Nanomaterials are particles, products, machines, or other engineered structures that are made on a scale between 1 and 100 nm. The process of designing, creating, and manufacturing products this small

is called “nanotechnology.” Sometimes the scale is extended to 1  $\mu\text{m}$  but technology on this scale is properly called microtechnology, not nanotechnology. Nanomaterials are very heterogeneous, including solid particles, complex structures that move with an energy source, and quantum dots (which are semiconductors with just a few thousand electrically active atoms apiece). They also include fluorophores (quantum dots that emit light), nanocrystals, buckyballs (spherical 60-carbon structures similar to a geodesic dome that may have metal ions or other constituents trapped inside), and other fullerenes (a class of structures, including buckeyballs, that follow the geometry of geodesic domes but may be extruded and elongated in the form of wires or nanotubes).

Nanotechnology deals with a world on the scale of bacteria and the larger viruses. Ultrafine particulate air pollution falls in this range, and nanoparticles, particularly fullerenes, are also natural products of combustion. Particles on this scale were not appreciated or even known until recent decades, and characterization of their properties has led to some surprises. The properties of these materials may be very different from the properties of bulk materials. For example, metals that are inert on a conventional scale, such as gold, may be catalytic on a nanoscale. Titanium particles are so opaque that the material is used in paints on a conventional scale, but on a nanoscale they are transparent to visible light and absorb ultraviolet radiation. Gold is its familiar color on a conventional scale but red on a nanoscale. One of the early nanomaterials to make it to the market is silver nanoparticles, which at that scale are so bactericidal that they may compromise wastewater treatment. Pure carbon particles on a conventional scale are dry, non-adherent powders, but pure carbon fullerenes are greasy.

Because of the great variety, complexity, and difference in properties among the various types of nanomaterials, it is not very productive to consider them as a class of materials, as done in this chapter and most contemporary sources. In the future, individual nanomaterials will have to be evaluated and considered as they are developed, as if they were new chemicals or chemical formulations.

For the moment, however, society is in a transition and is dealing with regulatory and risk assessment regimes that treat nanomaterials as a single class of materials, which they are not except in terms of size.

One of the major issues in risk assessment of nanomaterials, such as a simple carbonaceous particle below 100 nm (the same range as ultrafine particulate air pollution), is how to measure exposure or dose. The weight of nanoparticles is negligible because of the skewed distribution of mass with size of particles (see Chapter 2): small particles carry very little mass but have disproportionate biological activity. The number of particles may be more relevant, particularly if an effect requires uptake of a particle by a macrophage or bacterial cell. For many applications, such as catalyzed chemical reactions or absorption of a nanopharmaceutical, total surface area of the particle suspension is most important. For a given mass of nanoparticles, the total surface area is disproportionately larger than for a conventionally sized particle of the same mass, due to the relationship between surface area and volume. Other chemical and physical characteristics (electrostatic charge, aggregation, shape, composition, substitutions on the surface) may be critical in some situations and not others. This has prompted endless debate and discussion over the appropriate metrics for testing nanomaterials.

Without a clear metric for evaluating exposure, it is difficult to describe the toxicological properties of nanomaterials. Some generalizations do seem to be emerging from “nanotoxicology,” however. Nanomaterials tend to evade biological host defense mechanisms. They remain suspended in the respiratory tract and thus bypass the usual particle clearance mechanisms. Nanoparticles currently marketed in sunscreens do not seem to be absorbed transdermally, but it is easy to imagine structures that might be. They may be distributed by the circulation if they enter the vascular space, which is more likely for inhaled and ingested nanoparticles. They can cross the blood-brain barrier. They may be small enough to evade immune surveillance and scavenging by the reticuloendothelial system (including macrophages).

Carcinogenicity has been a concern for nanomaterials as for any new chemical product, but to date the risk does not appear to be disproportionate to that of other chemical substances. Preliminary studies suggest that nanotubes are capable of inducing mesothelioma when implanted in the pleura, but so are cotton fibers.

Risk assessment and risk management of nanomaterials is far from settled. In the end, it may be necessary to evaluate each product on its own by whatever method seems most appropriate to its application and properties. This opens the possibility of inconsistent risk assessment, outside the current template. Until this is resolved, it seems inevitable that conservative (i.e., more protective) assumptions will need to be made, and some investigators have proposed that a new “uncertainty factor” (see Chapter 7) should be introduced for standards setting for materials on a nano scale.

## OXIDES OF NITROGEN

The oxides of nitrogen ( $\text{NO}_x$ , often pronounced “knocks” in air pollution circles) are a large and complex family of compounds. The higher  $\text{NO}_x$  are those compounds of nitrogen at higher valence, such as the form emitted from automobile tailpipes, nitric oxide ( $\text{NO}$ ), and the secondary pollutant formed by photo-oxidation, nitrogen dioxide ( $\text{NO}_2$ ), and its brown dimer, dinitrogen tetroxide ( $\text{N}_2\text{O}_4$ ). The lower oxides of nitrogen, mostly nitrous oxide ( $\text{N}_2\text{O}$ ), an anesthetic and stratospheric ozone-depleting gas, are of little interest in occupational and environmental toxicology and will not be discussed further here. The higher oxides of nitrogen, on the other hand, are common and important hazards in occupational medicine and are major constituents of urban air pollution.  $\text{NO}_2$  is the most important as a toxic hazard and will be the principal chemical species discussed. In air pollution toxicology, “ $\text{NO}_x$ ” refers to the sum of  $\text{NO}$  and  $\text{NO}_2$ .  $\text{NO}_2$  is a simple compound with a “bent end-on” configuration, similar to carbon dioxide. Under usual ambient conditions, some 30 percent or so is present as the dimer, dinitrogen tetroxide ( $\text{N}_2\text{O}_4$ ).  $\text{N}_2\text{O}_4$  is responsible for the characteristic brownish fumes over nitric acid and

the brownish haze of photochemical smog. The  $\text{NO}_x$  are interconverted by a variety of reactions, of which the most pertinent to air pollution chemistry is the conversion of NO, a product of the internal combustion engine, to  $\text{NO}_2$ , and the subsequent dimerization of  $\text{NO}_2$  to  $\text{N}_2\text{O}_4$ . The conversion of NO to  $\text{NO}_2$  is relatively slow and can easily be observed over a period of hours after morning rush hour in cities such as Los Angeles.

The photochemistry of oxidant air pollution is exceedingly complex. The  $\text{NO}_x$  play key roles in the principal reactions:

- As end products themselves
- As sources of free oxygen radical
- As electron acceptors in the reduction of ozone to  $\text{O}_2$
- When atmospheric NO is depleted,  $\text{O}_3$  accumulates
- As reactants in the formation of peroxyacetyl nitrates (PAN)

The  $\text{NO}_x$ , and  $\text{NO}_2$  in particular, participate in a variety of chemical reactions of biological significance, which partly explains their toxicity. These include:

- Hydration of  $\text{NO}_2$  to nitric acid, which can occur on the surface of airborne particulates
- Direct oxidation of unsaturated fatty acids
- Free radical formation, producing very stable species when, for example,  $\text{NO}_2$  oxidizes phosphatidyl choline (lecithin)
- Nitrosation of amines to form nitrosamines
- Paramagnetic interactions between NO and cytochromes and hemoglobin, which has provided a useful tool in bio-physical research and which may have toxicological significance

As products of high-temperature combustion,  $\text{NO}_x$  are intimately associated with technological development. The hazard of exposure

to  $\text{NO}_x$  has been a mirror of Western social history and a cause of practical concern for 176 years.

The first documented case of human exposure to the  $\text{NO}_x$  occurred in 1804 in the shop of a French merchant in Lyon, who accepted a shipment of nitric acid in clay crocks, two of which were cracked and leaked vapors into the merchant's storeroom. Through the first half of the nineteenth century the peculiar natural history of  $\text{NO}_2$  exposure, in which the victim may feel little discomfort immediately after the episode but may die later of pulmonary edema, was described often in the medical literature.

In 1855 Alfred Nobel invented the blasting cap, which made the use of nitroglycerin and dynamite practical. Large quantities of nitric acid came to be used in the munitions industry, and in Germany alone hundreds of cases of toxic  $\text{NO}_2$  exposure were reported. After the First World War, human exposure to the  $\text{NO}_x$  was a hazard primarily in welding, mining, and the chemical industry. These and agriculture remain the principal sources of occupational risk today.

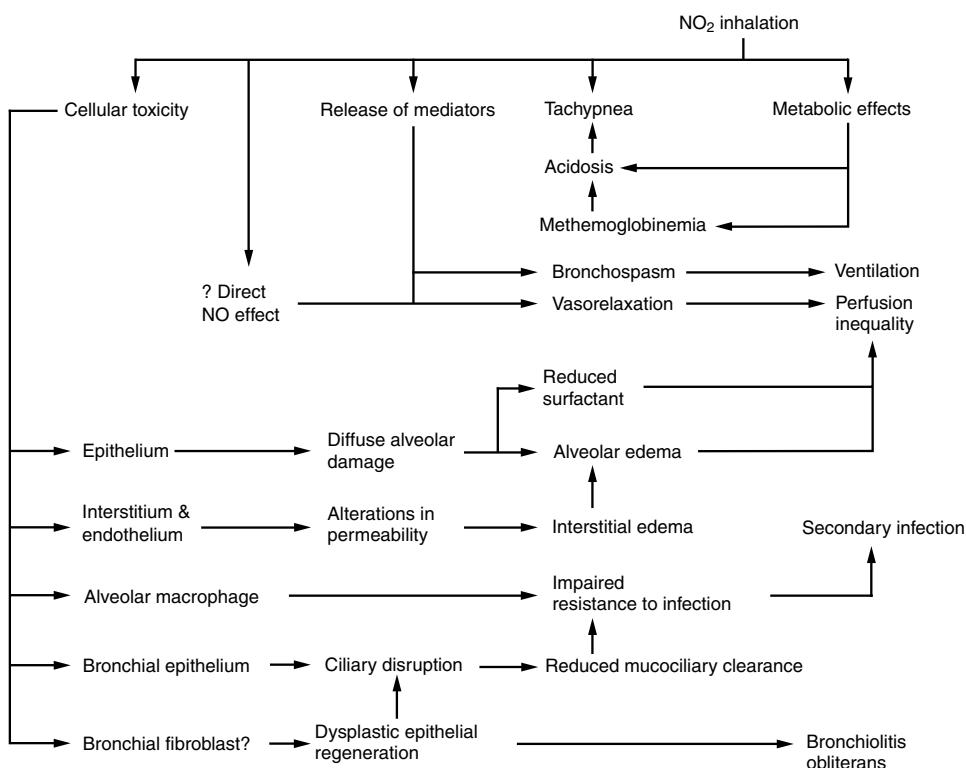
Silo filler's disease is a potentially fatal hazard of agricultural workers who enter a silo a few days or weeks after grains, especially corn, have been stored. The problem was not described until 1914, and for many decades the pattern of occasional sudden death and more commonly delayed respiratory distress was thought to be due to asphyxiation by  $\text{CO}_2$  and depletion of  $\text{O}_2$ .  $\text{NO}_2$  was recognized as the cause of silo filler's disease in 1950, however, when the oxidation of plant proteins into amines, and ultimately  $\text{NO}_2$ , was documented. Silo filler's disease is now one of the best-studied occupational lung diseases.

Accidental inhalation of high concentrations of the  $\text{NO}_x$ , especially  $\text{NO}_2$ , is a hazard in several industries and one major trade, welding. It is a hazard for welders and cutters in enclosed spaces such as the inside of a tank or a boiler. There have been releases of  $\text{NO}_2$  during upset conditions or leaks in chemical plants. Miners in the past have been exposed when they entered a blast area too soon after the explosion when a misfire had occurred. Agricultural workers still occasionally are at risk for silo filler's disease. The  $\text{NO}_x$  are also associated with high-temperature combustion.  $\text{NO}_x$  was the lethal

exposure for many of the victims in the Cleveland Clinic fire of 1929, which started with burning nitrocellulose x-ray films.

$\text{NO}_x$  is a relatively insoluble gas that is ineffectively cleared in the upper airway. Thus, it penetrates to the deep lung easily and damages bronchioles and alveoli.  $\text{NO}_x$  follows Haber's Law relatively closely. (See Chapter 2.) Over a broad range, the time-concentration product predicts toxicity, both for mortality and for pulmonary edema, at high concentrations.

The clinical syndrome (toxicodrome) is toxic inhalation and diffuse alveolar injury. The progression of the syndrome follows the process of pathophysiology outlined in Figure 10.4. First, the victim feels short of breath and may experience chest pain because of a direct



**Figure 10.4.** Pathophysiology of toxic inhalation due to nitrogen dioxide inhalation.

response of airways. Occasionally there will be nausea or overwhelming weakness and the victim may collapse on the spot. More often, he or she (usually a male) will get out of the gas cloud or will be removed by a rescuer. This acute reaction is caused by severe irritation to the bronchi and by constriction of bronchial smooth muscle, similar to that which occurs in an asthma attack. The acute reaction may look very much like a heart attack. Once the victim is out of the exposure situation, he usually recovers rapidly and sometimes goes right back to work. When the gas penetrates to the air spaces it causes a generalized pattern of injury to the air space walls, which is called "diffuse alveolar damage." About 12 hours after the exposure, often while they sleep, many victims experience a slow return of shortness of breath. Like the merchant in Lyon, they may die of pulmonary edema. Simply put, the vascular bed of the lung has been severely injured and has lost the integrity required to keep fluid from leaking into the air spaces, so the patient literally drowns. This process takes time as plasma accumulates in the wall of the air space (the interstitium) and finally passes into the air space itself. Thus, there is a delay of hours after the injury during which the patient may feel quite well and may be sent home from the emergency room.

Finally, the victim may be subjected to a third threat. After two or three weeks, the victim may become short of breath yet again. This time, the irritated lining of the bronchial tree may have repaired itself abnormally, scarring smaller airways and eventually obliterating them completely. This is called "bronchiolitis obliterans," and once established it is irreversible. These stages may be present singly or in any combination after an acute exposure. Additionally, the patient may develop a superimposed pneumonia (see Chapter 12).

The injury to the alveolar wall may be so severe that the alveolar wall loses its structural integrity. This results in a pattern of centrilobular emphysema, which is destruction of alveoli and an obstruction to airflow. Usually in combination with this pattern of emphysema, fibrosis occurs in the interstitium, leading to "honeycombing" and restrictive changes. The fibrosis results in a scarring in the airspace walls that makes the lung abnormally stiff.

Although the pathophysiology of acute NO<sub>2</sub>-induced injury to the lung has become a valuable animal model for diffuse alveolar disease, the chronic effects of exposure to low concentrations of the NO<sub>x</sub> remain largely unknown. It is known that NO<sub>2</sub> at levels near environmental exposure increases the seeding of cancer metastases in the lung in experimental studies. There are also strong suggestions that NO<sub>2</sub> may impair bodily defenses against infection.

In the late 1960s, the EPA organized a prospective study of respiratory tract infections and pulmonary function in 1,906 white non-asthmatic school children in Chattanooga, Tennessee. There, a local factory that manufactured dynamite was a point source of NO<sub>2</sub> emissions, and the children inhabited a neighborhood with high ambient NO<sub>2</sub> levels. The older children performed spirometry every week in November and March of two consecutive school years, and parents and family doctors of all of the children mailed in a weekly report on respiratory symptoms.

Although the design of the study appears to be simple enough, the investigators ran into a number of problems illustrating the difficulties of environmental epidemiology: The method used to monitor ambient NO<sub>2</sub> was the federal reference method, which in 1968 and 1969 was the Jacob-Hochheiser technique. This technique was proven worthless in 1972, invalidating all of the study's monitoring data. Fortunately, an old-fashioned but reliable "Saltzman" monitoring device had been set up on the outskirts of the high-NO<sub>2</sub> community on an Army post, so that at least some usable data were available. The neighborhood with the highest NO<sub>2</sub> level unexpectedly had the highest particulate level as well, but one neighborhood was intermediate in NO<sub>2</sub> and low in particulates and thus provided a positive control for particulates. The study did not monitor NO<sub>2</sub> indoors, and this is a significant omission. The study demonstrated that the children who lived in the neighborhoods with higher ambient NO<sub>2</sub> levels experienced more frequent lower respiratory tract infections (specifically bronchitis and croup) and were more likely to have repeated episodes of lower respiratory tract infections. These differences were statistically significant below  $p = 0.05$  and appeared to show increasing frequency of illness

with length of residence in the higher- $\text{NO}_2$  neighborhood. During the influenza epidemic of 1968, the attack rate for children in the higher- $\text{NO}_2$  neighborhoods was significantly higher than for those in the low- $\text{NO}_2$  neighborhood. Controlling for sex and month of test, the average ventilatory function of children ages 7 and 8 as measured by the  $\text{FEV}_{0.75}$  was slightly but significantly reduced among residents of the high- $\text{NO}_2$  neighborhoods compared to the lower- $\text{NO}_2$  ones.

The Chattanooga study, for all of its problems, suggests increased susceptibility to infection among children chronically exposed to low levels of  $\text{NO}_2$ , an exposure-response relationship with duration of exposure, and sufficient cumulative effect to slightly reduce pulmonary function in children.

Since then, laboratory findings have pointed in the same direction. There is now increasing evidence for several effects of  $\text{NO}_2$  on defense mechanisms of the respiratory tract.  $\text{NO}_2$  profoundly affects mucociliary transport at higher concentrations, but in much lower concentrations (2.0 ppm)  $\text{NO}_2$  potentiates the effects of influenza virus in slowing ciliary motility. Data also suggest changes in antibody titer after low-dose  $\text{NO}_2$  exposure, but these are not entirely consistent and appear to reflect species and strain differences.

Macrophages exposed to 15 ppm  $\text{NO}_2$  in vitro show markedly reduced phagocytosis, reduced bacteriocidal activity, and reduced myeloperoxidase activity, but preserved lysozyme and hydroxylase activity. At lesser concentrations, the depression in bacteriocidal activity is present at least down to 1 ppm. The macrophage is also the principal source of interferon, which is the body's primary defense against viral infection. The significance of these findings must be applied to the intact organism and its ability to resist infection. Here the animal data are most convincing. Effects have been shown in a number of primate and rodent species suggesting effects similar to those cited by the Chattanooga and other studies.

Since then, other issues in the regulatory process have diverted attention away from the  $\text{NO}_x$ , but there is growing sentiment for a review of both occupational and environmental standards. In the case of the occupational standard, concern is focused particularly on diesel

emissions, especially in underground mining. Interest in a review of the air quality standard is based on new evidence for more subtle effects of the  $\text{NO}_x$  at low concentrations. These include suggestions of an effect on immune function and a facilitative effect on metastases from existing cancers in animal studies.

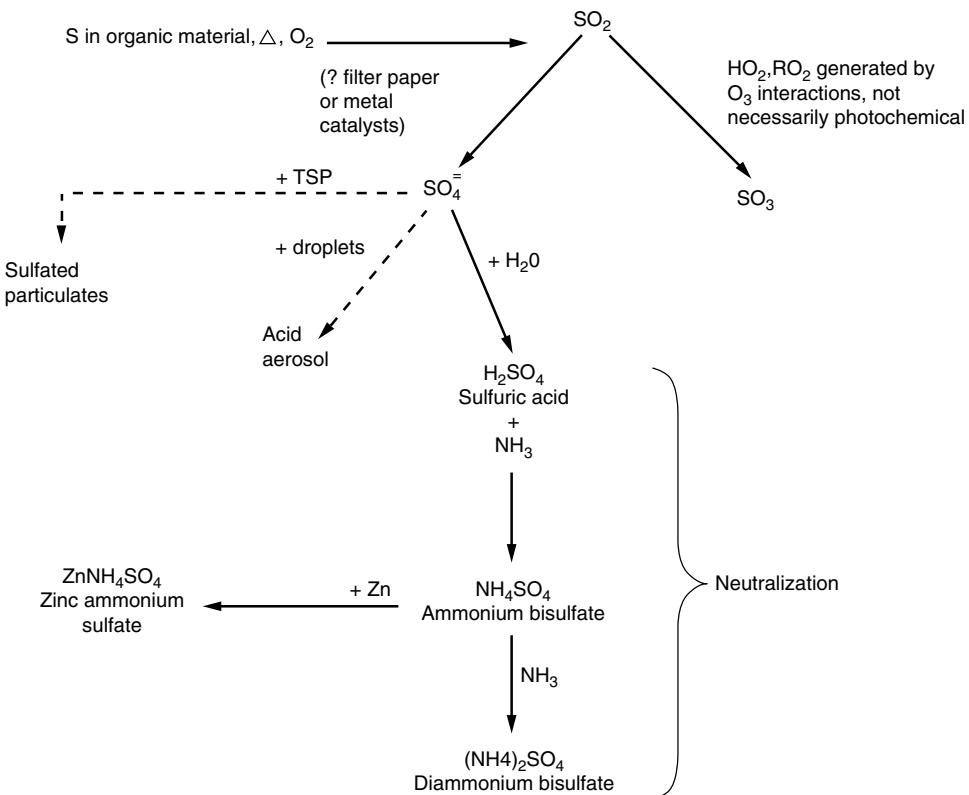
There is no evidence suggesting a role for  $\text{NO}_x$  as a primary carcinogen or classical promoter.

## OXIDES OF SULFUR

The oxides of sulfur ( $\text{SO}_x$ , often pronounced “socks” in air pollution circles) are critical constituents of the reducing variety of air pollution (see Chapter 12) formerly typical of heavily industrialized cities, historically in eastern North American and Northern Europe and more recently in China. This discussion will emphasize the atmospheric chemistry of the  $\text{SO}_x$  and describe effects of the  $\text{SO}_x$ , as occupational physicians are often concerned with the environmental issues pertaining to the  $\text{SO}_x$  in the course of their responsibilities. Figure 10.5 describes the interrelationships among the various oxides of sulfur.

$\text{SO}_2$  is highly soluble and consequently is absorbed readily across the mucosa of the upper respiratory tract. There is very little penetration of gaseous  $\text{SO}_2$  to the alveolar or terminal bronchial level unless a subject is exposed to overwhelming concentrations. Hydration of ambient  $\text{SO}_2$  in the upper respiratory tract may lead to intense local pH effects, but a significant amount of generated sulfate is neutralized by endogenous and mouth flora-generated  $\text{NH}_3$  at low ambient concentrations. The favorable chemical characteristics of  $\text{SO}_2$  are limited to the gas, however. In real exposure situations,  $\text{SO}_2$  is almost always produced in the presence of aerosols, usually particulate organic matter. The  $\text{SO}_2$  adsorbs onto the surface of the particulates as a thin film and is carried deep into the smaller airways and to the alveoli on particulates in the respirable range.

In humans, bronchoconstriction occurs at approximately 1 ppm  $\text{SO}_2$ . The degree of the response, measured as serial change in airway



**Figure 10.5.** Interrelationship of oxides of sulfur.

resistance for each subject, also defines a small group of apparently health non-asthmatic individuals.

In the middle 1960s, additional observations led to a substantial rethinking of  $\text{SO}_2$  toxicology. The toxicity of  $\text{SO}_2$  in combination with aerosols was found to be extremely sensitive to aerosol size, and the “ $\text{SO}_2 +$  particulate complex” was found to be much more hazardous than exposure to  $\text{SO}_2$  alone. This was discovered in the first large-scale population studies available linking  $\text{SO}_2$  and population health effects. These relationships had previously been difficult to assess because of limited monitoring technology and the problem of accounting for both total suspended particulates (TSP) and  $\text{SO}_{\text{x}}$ .

**Table 10.6.** Factors Altering Adsorption of SO<sub>2</sub> onto Particulate Matter

SO <sub>2</sub> concentration	(relationship is not stoichiometric)
Particulate concentration	
Humidity	(favors rapid uptake)
Organic content of particle	(favors reversible association)
Temperature	(200°, not significant at ambient temperatures)

Table 10.6 lists the factors known to influence the absorption of SO<sub>x</sub> onto particles.

The combination of SO<sub>x</sub> and particulates has been implicated in most of the serious air pollution incidents that have involved heavy mortality (see Chapter 12). Early air pollution episodes, such as Donora and the Meuse Valley, were usually related to a combination of atmospheric inversion, high SO<sub>2</sub> emissions, and high particulate levels. The era of British air pollution disasters largely ended with London's control of particulate emissions, despite this neglect of control of SO<sub>2</sub>. Smoke abatement alone helped, although in 1962 SO<sub>2</sub> levels equaled 1952 levels (1.5 ppm).

Mortality studies associated with fluctuating daily levels of SO<sub>x</sub> and TSP have shown effects associated with high levels in air pollution. The best data are from New York City, but numerous technical problems have interfered with interpretation. Using a time-series analysis technique, investigators have concluded that the influence of SO<sub>2</sub> on mortality in New York City in the 1970s was weak or non-existent, but that the influence of TSP on mortality was small but significant, predominantly in older groups earlier in the study period.

Morbidity studies have faced even greater methodological problems. Longitudinal observations on morbidity performed in London during the 1960s suggested that respiratory illness was associated with TSP more than with SO<sub>2</sub>. Other studies have suggested that the combined effect of SO<sub>2</sub> is particularly associated with cough, at a threshold of 0.17 ppm SO<sub>2</sub> and 0.25 mg/m<sup>3</sup> smoke for complaints of respiratory

symptoms. Among persons who already have bronchitis, these changes are highly confused by associated changes in other, weaker pollutants. Asthmatics show an increased frequency of attacks with exposure to SO<sub>2</sub> and TSP. Recent evidence suggests that asthmatics also react to sulfates in the absence of TSP and that this may be an effect due to long-range transport of acid substances in the atmosphere.

With control of both SO<sub>x</sub> and TSP by engineering at the source, the role of the SO<sub>x</sub> + TSP complex in air pollution is not as great in the developed world as it once was. The complex remains a difficult problem in the developing world and in rare cases of point source emissions in industry, usually due to fire or misadventure.

## PESTICIDES

Pesticides are chemicals that selectively kill or control the growth of organisms, other than microbial pathogens, that are associated with health risks or economic loss. The most important legal definition of a pesticide is drawn from the U.S. Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), which defines a pesticide as any substance or mixture of substances intended for preventing, destroying, repelling, or mitigating any pest, or intended for use as a plant regulator or to control plant growth. Pesticides are used in abundance in modern society for preserving and increasing agricultural production, controlling the vectors of human and veterinary disease, gardening, controlling aesthetic nuisances around the home and in public spaces, preventing structural damage to houses and wooden buildings, and keeping ground free of unwanted vegetation for reasons of safety, fire control, and commercial usage.

### Overview

Occupational exposure to pesticides occurs most often in agriculture but may occur in virtually any industry. Environmental exposure may also occur at higher levels through agriculture or through lawn and park maintenance and household use. Food may be contaminated as residues or at toxic levels, as has occurred with aldicarb. These products

are used in many settings other than the production of food and fiber: to maintain grounds or to reduce spoilage by insects or other biological pests. Significant exposure may occur in the home, in conjunction with controlling nuisance pests. A major problem has emerged with poorly supervised or regulated pesticide use in developing countries, with an alarming increase in pesticide-related toxicity incidents among workers. Since a much higher proportion of residents is engaged in agriculture in developing countries than in the United States or Canada, a much greater proportion of the population is at risk from higher-level exposure to pesticides in these countries.

There are broad trends in pesticide use that affect the frequency of toxic events and community levels of exposure. Because of concerns over bio-persistence, most of the persistent organochlorine pesticides have been banned, phased out, or removed by manufacturers from the market since the 1970s. They were initially replaced by the much more acutely toxic organophosphate and carbamate pesticides, which had the advantage of being readily degraded in the environment. These pesticides presented a serious problem of occupational health risk and in some cases food contamination. Over time, these acutely toxic pesticides are in turn being replaced by pyrethroids and other substitutes.

Occupational exposure to pesticides occurs most often in particular settings, or when certain social factors are operating, especially in the United States, where immigrant labor is essential for agricultural productivity. The exposure is more likely to occur when pesticides are handled casually, are mislabeled, or are repackaged from their original containers. It is much more likely to happen in agricultural settings when the workers are not fluent in the dominant language. Today, whether immigrants have legal status or are undocumented, many agricultural workers speak Spanish or another tongue as their native language, have a halting command of English, and are sometimes illiterate in their native language as well as in English. Usually there is little time spent in one-on-one instruction on safe handling of pesticides, resulting in careless handling and needless exposure. Instruction and training in third languages is even less common, despite the increasing diversity of agricultural workers in the United States.

Although still unsatisfactory in many (perhaps most) areas, the situation in the fields has gotten better, especially in California, and should be dramatically improved with the introduction of proposed OSHA regulations for field sanitation. In the absence of adequate facilities in the field for sanitation, workers have not been able to wash their hands, change clothes, eat without contamination, or attend to their basic needs.

Additionally, agricultural application of pesticides may affect all members of a family. In day-to-day operations of a family farm, even children may be involved in farm chores. Families often live in camps close to the fields. Exposure may also occur indirectly in the home through passive exposure, when members return from pesticide application without changing clothes. Small children may ingest potent pesticides or become inadvertently exposed to them on farms or even around the house in non-farm homes. Pesticides are also among the potently toxic chemicals most readily available for suicide attempts.

In the home, pesticide preparations are usually purchased ready to use. However, in agriculture, pesticides are usually bought as concentrates or powdered preparations for bulk use. They must be diluted with a vehicle, usually an oil (petroleum distillates) or water, before application. Application may be by hand, by motorized apparatus, or by aerial spraying. Dilution and application are the stages in pesticide use most likely to result in occupational exposure, either because of inadvertent contact with the mixture during the process or as the result of a spill.

The pesticide formulations usually contain several chemicals other than the active ingredient. These are called "inert" ingredients only because they are inactive with respect to pest control. However, these compounds are often far from inert chemically or toxicologically. They may include the vehicle, other pesticides, minor contaminants, and flow agents, which are chemical additives that keep the powders from caking. In cases that are confusing or in which the syndrome (toxicidrome) of pesticide exposure is not characteristic, the potential toxicity of the "inert" ingredients should be considered.

The most heavily used pesticides in North America are insecticides, herbicides, and fungicides. The chemical classes of heaviest use are the

organophosphates, carbamates, organochlorines, phenoxyacetates, and pyrethroids. The most common pesticides as causes of toxicity (omitting disinfectants, which are regulated as pesticides under the Federal Insecticide, Fungicide, and Rodenticide Act) are the organophosphate insecticides (among adults, particularly parathion), pyrethroids (which are common causes of toxicity in adults and children, usually low-grade), carbamates, organochlorines, and phenoxyherbicides.

Table 10.7 is a general guide to the acute management of pesticide toxicity prior to arrival at the hospital. It can be reproduced and distributed for the education of workers handling pesticides in the field. General measures should include skin decontamination (with soap and lots of water) and removal of contaminated clothing. Activated charcoal is used for ingestion, which is usually due to a suicide attempt. Care should be taken in such cases because the emesis material may be hazardous.

It is useful to ask those who attended the patient to bring in any containers of pesticide or other work-related chemicals the patient may have been using, for inspection. Product names can be misleading and formulations may change. It is wise to use the generic name for the pesticide in looking up the toxicity of the active ingredient. If the pesticide container is available in a case where the toxic agent is unknown, the active ingredient should be specified on the label and in a material safety data sheet for the product. A standardized vocabulary is used by EPA, with the following meanings:

- “Danger” means a product that is highly toxic, corrosive, may cause irreversible eye and skin damage, and can be associated with fatal toxicity or allergy.
- “Warning” means a product that is moderately toxic, may cause moderate injury and irritation, and for which certain precautions should be taken: avoid contact, remove contaminated clothing, and wash thoroughly.
- “Caution” means a product that may be harmful and with which contact should therefore be avoided.

**Table 10.7.** General Guide to Managing Pesticide Toxicity Prior to Arrival at Hospital

**FIRST AID FOR PESTICIDE POISONING**

General symptoms of pesticide poisoning include: headache, giddiness, nausea, blurred vision, chest pains, sweating, drooling, vomiting, and contracted pupils.

- DO NOT PLACE YOURSELF IN DANGER TO RESCUE A POISONING VICTIM. Remove the victim from the contaminated site to a fresh air location, wearing a suitable respirator if fumes are present, and wear pesticide-impermeable gloves and coveralls to prevent skin absorption. In emergency situations where pesticide is in the air, self-contained breathing apparatus (SCBA) is required.
- IF THE VICTIM IS NOT BREATHING: Administer mouth-to-mouth artificial respiration until adequate breathing has been restored. Be sure to clean any contamination from the victim's face before administering artificial respiration. Call an ambulance or take the victim to the nearest hospital emergency ward.
- IF THE VICTIM IS BREATHING BUT UNCONSCIOUS: Remove contaminated clothing, taking care to avoid contaminating yourself and wash contaminated skin (see note below). Position the victim in a semi-prone position turned slightly to one side. This will prevent choking should vomiting occur. DO NOT GIVE ANYTHING BY MOUTH TO AN UNCONSCIOUS PERSON. Call an ambulance or take the victim to the nearest hospital emergency ward.
- IF THE VICTIM IS BREATHING AND CONSCIOUS: Remove contaminated clothing and wash contaminated skin thoroughly with soap and water. (If possible, skin should be washed with continuous running water for at least 10 minutes).

Find the pesticide label and bring it to the telephone. Phone the nearest poison control center and provide the staff with: an assessment of victim's breathing; the exact name of the pesticide as it appears on the product label; circumstances of the exposure (fumes, spray drift, spill, etc.); symptoms; what action has been taken; the name and age of the victim; and the phone number of where you are calling from or where you are taking the victim to.

Do not give medications or induce vomiting before calling the poison control center. Never use salt water to induce vomiting. Do not

*(Continued)*

**Table 10.7.** (Continued)

---

induce vomiting unless you are advised to do so by the poison control center.

If you are unable to contact the poison control center, follow first aid instructions on the pesticide label and contact the center as soon as possible.

- DO NOT INDUCE VOMITING UNLESS RECOMMENDED BY THE POISON CONTROL CENTER.
  - IF A PESTICIDE HAS BEEN SPLASHED INTO THE EYES: Immediately bathe the eye in a gentle stream of clean water and continue for at least 15 minutes. Report to the nearest hospital emergency ward.
  - PREVENT PESTICIDE POISONING
    1. Always read the label carefully before using any pesticide product.
    2. Store pesticides in locked storage areas out of the reach of children.
    3. Store pesticides in their original containers to avoid confusion with other products.
    4. Always wear appropriate protective equipment when handling and applying pesticides, including respirators, impermeable gloves and aprons, and goggles.
    5. Change clothing worn to apply pesticides at least once per day and launder thoroughly before re-using.
    6. Wash thoroughly with soap and water after handling or applying any pesticide.
    7. Always mix pesticides outdoors in a well-ventilated area.
  - FIRST AID SUPPLIES TO KEEP ON HAND (Checklist)
    - ✓ Clean water, soap, and towel for accidental spills. If you will not be close to washing facilities, be sure to carry these supplies with you.
    - ✓ Syrup of ipecac (available at most drug stores) to induce vomiting if required. Do not use ipecac unless instructed to do so by the poison control center or a physician.
    - ✓ Activated charcoal (available at drug stores) to absorb pesticides that may have been ingested. This product should only be used as instructed by the poison control center or a physician.
    - ✓ Bandages and tape to cover minor cuts that may facilitate pesticide entry through the skin. Waterproof gloves should always be worn when handling pesticides. Latex gloves are not protective.
    - ✓ Coins for emergency telephone calls.
-

Every pesticide is given an EPA Registration Number with eleven digits (the first four are unique to the company, the second three to the product, and the last four to the distributor). This number can be used to access toxicity information on Poisindex, which is available to any poison control center.

As a practical matter, pesticides must be discussed by chemical class. There are too many pesticides to deal with individually, and the clinical characteristics of acute toxicity by members of the most important classes of pesticides, the organochlorines, organophosphates, carbamates, and phenoxyacetic acid herbicides, are very similar within classes.

Pesticides may be classified by their function or by their chemical structures. A mixed system is usually the most satisfactory, if not the most tidy intellectually. A classification by chemical type is presented in Table 10.8 for the major chemical categories of pesticides, as judged by volume of application and frequency of use. Although toxicologists frequently refer to pesticides by chemical group and usually use their generic names in professional communications, workers in agriculture generally refer to these agents by their functional classification or brand name.

### **Organophosphate and Carbamate Insecticides**

Both chemically and toxicologically, the organophosphate and the carbamate pesticides are similar classes (see Figure 10.6). They overlap in potency, but the majority of organophosphates are generally more toxic than most carbamates. The organophosphates and the carbamates are closely related chemical classes chiefly used as insecticides. Their bioavailability is high and absorption is very efficient. Distribution is by perfusion, and some crosses the blood-brain barrier, but there is no appreciable storage depot. Metabolism to paraoxon is rapid and produces the active agent. Elimination is mostly by metabolism.

Organophosphates and carbamates act by the same mechanism as neurotoxins in both insects and humans, producing a stereotyped

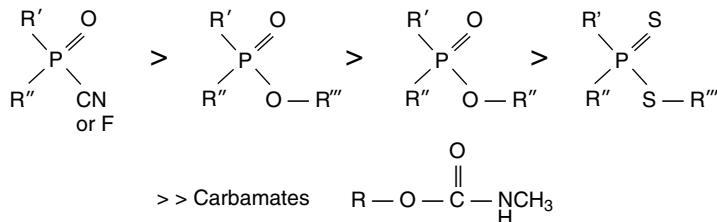
**Table 10.8.** Major Chemical Classes of Common Pesticides

Classification	Common Uses	
	Examples	Representative
<i>Organochlorine:</i>		
Halogenated biphenyls*	Insecticides	DDT, methoxychlor
Cyclodienes*	Insecticides	Chlordane, aldrin
Chlorohexanes*	Insecticides	Benzene hexachloride (lindane), chlordcone, mirex, toxaphene
Halogenated aliphatics	Fumigants Nematocides	EDB, EDC, DBCP
Chlorophenol preservatives	Wood	Pentachlorophenol, tetrachlorophenol
Fungicides		Triclosan
Phenoxyacetics	Herbicides	2,4-D, dicamba, dichlorprop
<i>Other</i>		
Organophosphates	Insecticides	Parathion, malathion, mevinphos, phorate, fonofos
Carbamates	Insecticides	Carbaryl, carbofuran, methomyl, propoxur, aldicarb
Pyrethroids	Insecticides	Pyrethrum, allethrin
Thiocarbamates	Herbicides Fungicides	Thiram, ziram
Bipyridyls	Herbicides	Paraquat, diquat
Aminotriazines	Herbicides	Aminotriazole, atrazine
Metallooxins	Fungicides	Copper salts, mercurials
Coumadin derivatives	Rodenticides	Warfarin, pindone, diphacinone
Indandiones	Rodenticides	

\*Usually grouped together for toxicological purposes as the “organochlorines.”

### ACETYLCHOLINESTERASE INHIBITORS

Structure —Activity relationship for mammalian toxicity



### ORGANOPHOSPHATE PESTICIDES

		$\text{LD}_{50}$ (Rat) mg/kg, po
TEPP	$\text{E}_{14}\text{P}=\text{O}$	1
Parathion	$(\text{C}_2\text{H}_5\text{O})_2 - \text{P} = \text{S} - \text{O} - \text{C}_6\text{H}_4 - \text{NO}_2$	13
Malathion	$\begin{array}{c} \text{CH}_3\text{O} \\   \\ \text{CH}_3\text{O} - \text{P} = \text{S} - \text{CH} - \text{COCH}_3\text{CH}_3 \\   \\ \text{CH}_2 - \text{COCH}_3\text{CH}_3 \\    \\ \text{O} \end{array}$	1375    15.25 g

### CARBAMATE PESTICIDES

Carbaryl	$\begin{array}{c} \text{O} \\ \parallel \\ \text{O} - \text{C} - \text{N} - \text{CH}_3 \\   \\ \text{H} \end{array}$	500–850
----------	---	---------

**Figure 10.6.** Organophosphate pesticides.

toxicity profile than makes it more convenient to discuss them together. Because insects have a lower tissue level of the hydrolyzing enzymes that deactivate the agents than do humans, they are much more susceptible to the toxicity of the agents.

The organophosphates are the leading cause of fatal toxicity involving pesticides in humans. The carbamates are closely related chemically and share most toxicological characteristics, but there are

significant differences in potency and duration of effect. In general, the carbamates are less potent as a class and the effect is much shorter, because, unlike the organophosphates, they do not bind irreversibly to the enzyme active site. This is only a general rule, however. TEPP, parathion, and methyl-parathion are very potent, more potent in terms of human toxicity than any carbamates. Aldicarb and carbofuran, however, are among the most potent carbamates and are more potent in humans than many organophosphates, including dichlorvos and iodofenphos. All of these are relatively more toxic than carbaryl (popularly known as Sevin®), which, although a carbamate, is in turn more toxic than malathion, an extremely weak organophosphate in terms of human toxicity, and diazinon, which is used for ground applications. While it is useful to identify the specific agent responsible for a toxic event in order to access the likely severity of the acute episode and whether or not aggressive treatment will be required, the principles of management are uniform among chemical classes except that pralidoxime (discussed below) is not used in cases of known carbamate toxicity.

The toxicity of other pesticides derives, in general, from the direct action of the agent or a metabolite. In the case of the organophosphate and carbamate insecticides, the toxicity results from the accumulation of an endogenous neurotransmitter, acetylcholine, due to inhibition of the enzyme responsible for lysing acetylcholine and freeing the postsynaptic receptor for further stimulation. Since the toxicity relates to the accumulation of neurotransmitter rather than to the agent directly, the effects of these two classes are very similar.

The oxon or carbamate group combines with a serine-containing active site on the acetylcholinesterase molecule where the acetyl moiety of acetylcholine (AcCh) fits. This occupies the site, resulting in non-competitive inhibition and an accumulation of AcCh that cannot then be hydrolyzed. The carbamate is spontaneously hydrolyzed and the previously carbamylated enzyme is soon regenerated, with renewed activity. The organophosphates, however demonstrate an “aging” process in which the bond becomes covalent

at some point, which permanently destroys the activity of the enzyme. Recovery then depends on production of new enzymes.

The symptoms of poisoning by either class can be categorized by their physiological characteristics within the autonomic nervous system. The autonomic nervous system is divided into two principal branches, the sympathetic and the parasympathetic, the latter of which regulates principal bodily functions on a more or less automatic basis. The parasympathetic nervous system is further divided into pathways with receptors that are characterized as "muscarinic" or "nicotinic" (depending on their response to pharmacologic agents). The muscarinic pathways are primarily concerned with glandular secretion, cardiac pacing, and the iris. The nicotinic pathways are concerned with muscle activity. The two classes of insecticide also have central effects, which complicates the clinical picture, and some sympathetic effects may be seen because of the stress associated with full-blown toxicity. The treatment of toxicity by acetylcholinesterase inhibitors is unusual in toxicology in having not only one specific antidote, but two: pralidoxime and atropine (effective for both classes of pesticides). Furthermore, there are specific clinical signs available, such as miosis, for titrating treatment and laboratory markers (cholinesterase levels) to assess the magnitude of toxicity.

Both the organophosphates and the carbamates act by inhibiting acetylcholinesterase, the enzyme that hydrolyzes AcCh in the central and peripheral nervous system, creating an accumulation of AcCh at the postsynaptic junction, or endplate. This results in sustained tonic, uninhibited stimulation of the parasympathetic nervous system and skeletal muscle contraction, with effects that are predominantly muscarinic, some nicotinic. The primary symptoms of organophosphate and carbamate toxicity, therefore, are those of unopposed parasympathetic activity, skeletal muscle dystonia, CNS effects (representing a breaching of the blood-brain barrier), and selected sympathetic effects as a secondary response. There are no symptoms unique to one class of AcCh-inhibiting pesticide or another.

The other class of parasympathetic symptoms are the nicotinic effects, which also act on the voluntary nervous system, affecting

skeletal muscle. The signs and symptoms associated with the nicotinic effects include fasciculation (obvious involuntary twitching), cramps, muscle weakness (sometimes progressing to paralysis), tachycardia, hypertension, and pallor. The muscle weakness may be persistent because of prolonged neuromuscular blockade after very heavy exposure. These effects are specifically treated or, in the case of late neuromuscular blockade, prevented by administration of pralidoxime when the cause is organophosphate toxicity. Carbamate toxicity is much more readily reversible and does not require this treatment.

In combination, these symptoms constitute the characteristic toxidrome for these pesticides, described by the unfortunate mnemonic “SLURGE” (salivation, lung secretion, urination, respiratory distress, GI symptoms, and excrement, referring to bowel mobility). The patient is often flushed and disoriented, and may be cyanotic. The diagnosis of, in particular, acute organophosphate toxicity is not always straightforward, especially at lower levels of toxicity. The presentation may resemble heat stroke, myocardial infarction, pneumonia, asthma, influenza, alcoholism, encephalopathy, epilepsy, and diabetic hyperglycemic coma. It may also be very mild, presenting only as headache, dizziness, and nausea, resembling influenza.

Pralidoxime (2-PAM, 2-pyridine-aldoxime methiodide) is a specific treatment for organophosphate toxicity and works particularly well for parathion, dichlorvos, and diazinon. To prevent confusion of laboratory results, cholinesterase levels should always be drawn before 2-PAM is given, not after. 2-PAM should be given early, however, in order to regenerate acetylcholinesterase before aging occurs. If there is a delay, it should still be given later, however, to prevent new inhibition of the regenerating enzyme by circulating free organophosphate. The effectiveness of treatment, and the dose given, is titrated to the symptoms. 2-PAM has to be given slowly because it is a partial agonist and can cause respiratory depression or neuromuscular blockade if rapidly infused. Normally, however, side effects are minimal.

Because carbamates are not as tightly bound, and the enzyme regenerates quickly without 2-PAM, it is not useful in cases of

carbamate toxicity. If the agent is not known, 2-PAM probably does not cause harm in such cases and should probably be tried. If it is known that a case is due to carbamate toxicity, however, 2-PAM is withheld because of the theoretical risk of contributing to toxicity, since 2-PAM is also a weak agonist.

2-PAM has unusual chemical and pharmacological properties: Although a highly positively charged quaternary ammonium compound, it crosses the blood-brain barrier easily. Usefully, it also has some intrinsic anti-muscarinic properties and directly inactivates some free circulating organophosphate molecules.

The other antidotal treatment for organophosphate toxicity, which is also used in carbamate toxicity, is atropine, the familiar acetylcholine antagonist. Atropine competitively antagonizes acetylcholine at the muscarinic receptor and has little effect centrally or at nicotinic receptors. Massive doses may be required in severe cases, especially to correct cyanosis. The atropine effect can be monitored by observing heart rate, miosis, and reduction in primary symptoms. Signs of over-atropinization, on the other hand, are dry skin, decreased bowel sounds, tachycardia, and mydriasis.

Supportive care is required for severe cases, including seizure control with diazepam. Cholinergic drugs and opiates are contraindicated.

Central effects of toxicity from these pesticides, almost always the organophosphates, may include headache, convulsions, coma, anxiety, a labile affect, confusion, delirium, ataxia, slurred speech, and respiratory failure. A toxic psychosis and sudden loss of consciousness may indicate previously unsuspected serious toxicity. These effects are treated by supportive care, and may persist after the somatic effects are under control. They do not respond to atropine or 2-PAM.

Hyperglycemia may be observed secondary to epinephrine release, adding a sympathetic effect. The bradycardia of the muscarinic effect and the tachycardia of the nicotinic effect may roughly cancel, giving a relatively normal rate (often paradoxically normal in the face of an agitated, toxic individual), but this is highly variable. If ingestion is the route of entry, the emesis material brought up by the patient must be handled carefully, with impermeable gloves and apron,

because it may contain high concentrations of the agent. Respiratory depression may be an effect of the organophosphates but could also be an effect of the petroleum vehicle of the pesticide in cases of ingestion. Sympathomimetics or theophyllin compounds should be avoided in treatment because of the risk of cardiac arrhythmias.

The diagnosis of organophosphate or carbamate toxicity is usually based on a history of exposure. In the clinical setting, the impression of organophosphate toxicity in particular can be confirmed by the response of the patient to a therapeutic trial of atropine. The diagnosis should be confirmed by cholinesterase determinations. Cholinesterase level determinations remain the principal means of confirming the diagnosis of organophosphate or carbamate toxicity.

As inhibitors of the acetylcholinesterase enzyme, the degree of toxicity of these agents can be measured by the depression from normal levels in the activity of the enzyme. Because both the organophosphates and the carbamates are non-selective with regard to which cholinesterase enzymes they inhibit, exposure to them can be monitored by measuring the activity of accessible blood cholinesterase levels that are affected as well. There are two cholinesterase forms available for routine monitoring: plasma and red cell cholinesterase. Plasma cholinesterase levels are labile, responding immediately to toxic exposure and recovering quickly; plasma cholinesterase activity is the sum of the activities of a mixture of cholinesterases in the plasma fraction of blood. Plasma cholinesterase levels reflect immediate past exposure. They are also acute phase reactants and may be incidentally elevated on screening tests. Red cell cholinesterase activity represents one enzyme, specific for acetyl-choline; it is closer biologically and, because it is intracellular, physiologically to acetylcholinesterase activity in the nervous system. Red cell cholinesterase levels change somewhat more slowly and remain markedly depressed for at least two weeks after heavy exposure, because enzyme activity is regenerated only with the production of new red blood cells. Red cell cholinesterase levels therefore reflect past exposure over a longer period and more closely reflect approximate tissue levels. Red cell cholinesterase levels are therefore preferred for

assessing the degree of toxicity or for evaluating chronic toxicity, but the test is not as readily available as plasma cholinesterase.

Both cholinesterase determinations have their uses clinically, and either can be obtained, depending on whether the exposure is very recent or has taken place hours or days before. Serial red cell cholinesterase determinations are also useful for following potentially exposed workers over time to assess apparent or subclinical exposure. It is best to have the worker serve as his or her own control, comparing values over time to the individual's own baseline. Individual cholinesterase levels in blood are very variable and relatively labile naturally. Comparison of individual values against a laboratory range of normal is not very helpful unless the cholinesterase levels are profoundly depressed below the range of normal. Occasionally, significant toxicity may occur while the individual's cholinesterase level remains within the normal range, in cases where the baseline level was unusually high and the depression was not quite below the lower limit of the normal range. It is therefore very useful to obtain baseline red cell or plasma cholinesterase determinations before an individual worker has the opportunity to become exposed, for use as an individual benchmark in case toxicity is later suspected. This is required by law or regulation for pesticide applicators and handlers in many jurisdictions as a routine monitoring procedure.

Cholinesterase levels should always be drawn before the administration of pralidoxime to avoid confusing the laboratory determination. Levels do not have to be repeated serially to follow the patient clinically; once at the beginning, before treatment, is enough to confirm the diagnosis. Because they require some time to perform, treatment should never be withheld pending return of the results; the response to atropine provides sufficient clinical confirmation to proceed. A small group of people have abnormally low levels of plasma cholinesterase activity on a hereditary basis; they can be followed by their red cell activity.

The organophosphates are also associated with a delayed neurotoxic effect, or organophosphate-induced delayed neuropathy (OPDIN),

first recognized among chemists who worked with these compounds, and documented in several industrial incidents involving leptophos. This syndrome may resemble Guillain-Barré syndrome or the motor paralysis seen in toxicity from other, non-pesticide organophosphates, particularly tri-ortho-cresyl phosphate (so-called ginger Jake paralysis). The diagnosis is made by exclusion, ruling out other causes of neuropathy, and by a history of exposure to an organophosphate known or suspected to cause this effect. It is distinct from the delayed or prolonged neuromuscular blockade that sometimes occurs in cases of heavy exposure to organophosphates and it cannot be prevented by early use of pralidoxime. The mechanism appears to be inhibition of another enzyme, a “neurotoxic esterase,” distinct from acetyl-cholinesterase. Currently in the United States, EPA pesticide screening regulations require a specific, sensitive test (using susceptible animal species such as birds pre-medicated with atropine) to identify and to exclude from the market new organophosphates that have the potential to cause OPDIN.

In the 1980s, a third or “intermediate” neurotoxicity syndrome (IMS) was described (by Senanayake and Karalliedde) in heavily exposed workers one to four days following exposure, during recovery from the acute cholinergic effects. The IMS occurs in approximately 20 percent of patients following oral exposure to OP pesticides, with no clear association between the particular OP pesticide involved and the development of the syndrome. This symptom complex involved rapidly progressive weakness of muscle enervated by cranial motor nerves, proximal limb muscles, respiratory muscles, and neck flexors. No fasciculations were observed; an unusual electromyographic pattern was observed. The symptoms do not respond well to atropine and oximes; however, patients survive with ventilatory support. Current research indicates that the extent of muscle weakness may vary from patient to patient, and some may not require ventilatory care; however, observation and close monitoring of respiratory function is necessary. A similarly timed neurotoxic syndrome was reported after massive ingestion of carbaryl, a carbamate insecticide.

Both organophosphates and carbamates have a short residence time in the environment, degrading quickly by hydrolysis. Occupational exposure usually occurs in settings of crop contamination or incidents of application. Occasionally they are ingested as food residues, but this is uncommon. The dry organophosphate powder is usually mixed in oil or, especially in the case of malathion, applied by air, with bait intended to attract a flying insect. The organophosphates usually have a pungent, disagreeable odor. The carbamates are more water soluble, although they are usually sold as a powder, in oil, or in methanol. Exposure in agricultural settings is usually transdermal, but inhalation also occurs and may be significant because of aerial spraying and the aerosolization and dispersion of the pesticide-containing oil droplets during application. The carbamates, especially, are common household pesticides. Ingestion of these pesticides may occur during suicide attempts or, accidentally, by children. Distribution is rapid once the pesticide enters the bloodstream. Toxicity results before there is appreciable storage. Metabolism of the organophosphates occurs peripherally by the action of the MFO system by the transulfuration reaction, substituting an oxygen for the double-bonded sulfur and thereby creating an oxon derivative, which is the active form. Deactivation occurs by hydrolysis, among other pathways. The carbamates are hydrolyzed but not activated by transulfuration; they can react with the active site of the acetylcholinesterase enzyme in their native form.

### **Organochlorine Insecticides**

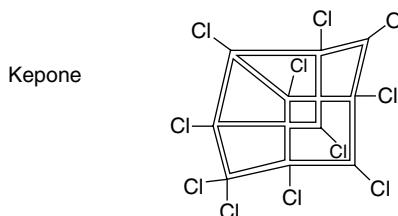
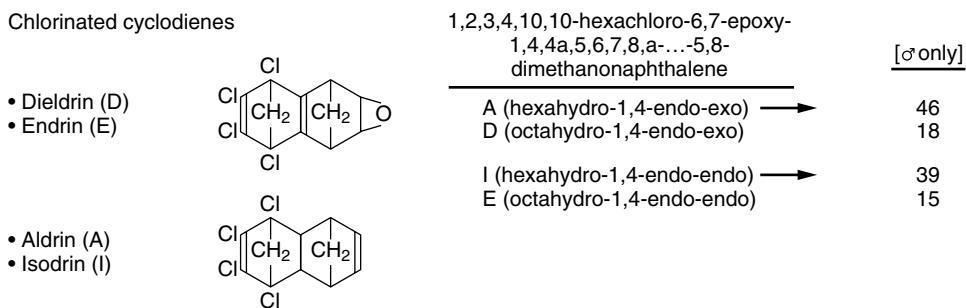
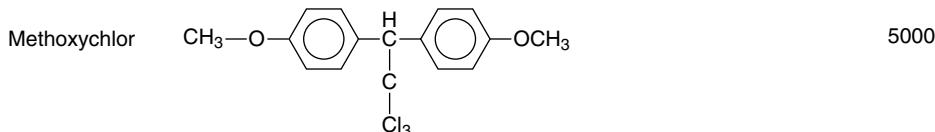
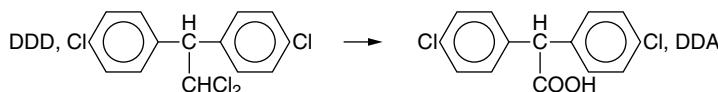
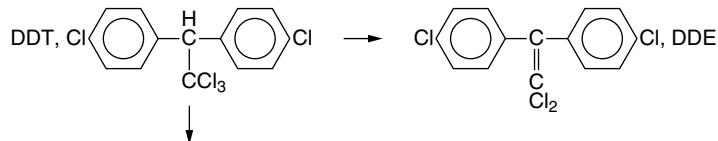
The organochlorines are a large and heterogeneous class of compounds (see Figure 10.7). Most organochlorines are not acutely toxic to humans except in massive exposures. The toxicity of the organochlorine pesticides varies widely, from chlordane and endrin, the most toxic, down through aldrin, dieldrin and isodrin, lindane, toxaphene, DDT, and methoxychlor, the least toxic of the series. All have been or are being phased out as persistent organic pollutants. Their characteristic toxicokinetic profiles vary considerably, from

## ORGANOHALIDE PESTICIDES

LD<sub>50</sub> (Rat)  
mg/kg, po

## DDT 1-p, p'-dichlorodiphenyl-2-trichloroethane

217

**Figure 10.7.** Organochlorine pesticides.

rapidly excreted agents such as lindane, methoxychlor, and dicofol (days), through intermediate agents such as the cyclodienes (weeks), to very persistent agents such as dichloro-diphenyltrichloroethane (DDT), chlordenecones (kepone and mirex), and the beta isomer of benzene hexachloride (months or years).

The absorption of the organochlorine pesticides is governed by their lipid solubility, which is generally very high, and that of the vehicle in which they are dissolved, usually a petroleum distillate or other oil. Transdermal absorption is considerable, and is the most common accidental route of exposure because of poor handling technique and failure to wear protective equipment. Inhalation is a relatively minor route of exposure for these agents. Ingestion is a minor problem, since the acute toxicity of this class of pesticides is not high and they are therefore unsuitable for suicidal attempts. Because of their solubility in lipid, these compounds tend to distribute into fatty tissue quickly and are released from their adipose stores only slowly. Blood levels are therefore misleading in reflecting body burden, but subcutaneous fat biopsies can be used to gain a more accurate estimate. Neither, however, are routinely obtained clinically. Metabolism relies heavily on the MFO system. Members of one group of the organochlorines, the cyclodienes, are capable of forming unusually stable epoxide derivatives with Phase I metabolism. This feature greatly increases their potential for toxicity, and members of this group have been found to be carcinogenic. Excretion is generally renal or hepatobiliary; certain members of the class, in particular the chlordenecones, have high rates of enterohepatic recirculation and are effectively conserved within the body by this mechanism. Chlordenecone toxicity has been treated using binding agents (cholestyramine) to interrupt re-absorption.

The organochlorines are neurotoxins designed to act selectively on the central nervous system of insects, inhibiting axonal transmission. In massive doses in human beings, the agents produce central effects, particularly anxiety, excitability, disorientation, behavioral changes, and sensory changes including paresthesias. These are associated with neuromuscular signs, including fasciculation, tremors, myoclonus,

weakness, and, rarely, a motor polyneuropathy. Seizure-like convulsions may occur; and in extreme toxicity, respiratory arrest may be the cause of death. The depressant effect of the oil-based vehicle is likely to be as or more significant than the primary effect of the pesticide in producing respiratory depression. In practice, fatal toxicity from the organochlorines is exceedingly rare. Hepatotoxicity is a common feature of the class, reflected in elevated enzyme patterns typical of hepatocellular injury. Myocardial irritability and the possibility of arrhythmias may result, partly on the basis of the actions of the agents but mostly on the epinephrine release in the clinical situation. Except for endrin, a particularly toxic pesticide, the prognosis for survival and full recovery following organochlorine intoxication is generally good. Residues will persist in the body, however, and in the case of nursing mothers these residues may expose the infant through breast milk.

Organochlorine toxicity is generally suggested by the history. There are few clinical signs that are characteristic other than the oculomotor clonus of kepone. Laboratory studies are usually not very helpful in assessing the level of toxicity or the prognosis. Abnormalities may be observed in liver and renal function tests. Residues are generally detectable in urine, blood, and fat by gas-liquid chromatography performed within seventy-two hours, but these findings are more useful forensically than clinically.

Treatment of toxicity by the organochlorines is primarily supportive. Ventilatory support, if needed, control of seizures with diazepam, and decontamination are the mainstays. Emesis should not be induced when the pesticide has been dissolved in a petroleum-based vehicle, because of the risk of aspiration and chemical pneumonitis. Oral sodium sulfate may hasten fecal evacuation and reduce absorption, but oil-based laxatives will increase absorption. Cholestyramine may hasten the excretion of agents that are re-circulated enterohepatically, such as kepone. Epinephrine and other cardiac stimulants should be withheld because of myocardial irritability.

Although the organochlorines as a class have relatively low acute toxicity, individual organochlorines may have specific actions that

present a greater hazard than that which exists for the class overall. It is therefore useful to identify the specific organochlorine pesticide responsible whenever possible, in order to anticipate the clinical course.

Chronic toxicity is a concern for all of the organochlorines, although few chronic effects have been thoroughly documented. DDT, kepone, DBCP, EDB, and toxaphene all have been categorized as group 2B carcinogens by the IARC. This indicates that these chemicals have been confirmed as animal carcinogens and show indications of being possible human carcinogens. Toxaphene is potently mutagenic without metabolic activation; this raises concern for its carcinogenic potential.

DDT, and its metabolite DDE, have low toxicity to human beings. However, because of their environmental persistence and effect on the metabolism of birds, they have been banned in most countries for most uses. It continues to be used, selectively, for malaria control. The history is unfortunate because very small applications of DDT were extremely effective in mosquito control largely because of this persistence, eradicating malaria in many regions after the Second World War with a single application to the walls of homes. The problem with ecotoxicity arose from its indiscriminate use in agriculture at a much greater volume and intensity. There is an ongoing debate over whether this minor remaining use should also be banned, but public health authorities generally favor its continued use because of its safety, effectiveness, the urgency of malaria control, and the very low amounts involved.

Chlordane was primarily used for fumigation and to treat foundations of houses for termite control before it was banned in 1988. There were many complaints of health effects due to heavy use of chlordane in homes, but they cannot be easily evaluated. Toxicity is nerve-related, associated with seizures.

Heptachlor was a restricted organochlorine that is allowed to be used against the fire ant, a pest that has been particularly severe in the southern United States. It is highly persistent and was the focus of Rachel Carson's seminal book *Silent Spring* (1962).

The “drin” pesticides (dieldrin, endrin, aldrin, and isodrin) were withdrawn in 1974 but are highly persistent.

Kepone was a chlordecone with a highly unusual structure. It was manufactured for a time in Hopewell, Virginia (up to 1975, when it was banned) in a famously poorly maintained plant that polluted the James River and, because of this incident, closed it to fishing. An outbreak of neurotoxicity among workers at the plant featured a unique form of ocular nystagmus. In some cases, the symptom started after the patient lost weight and mobilized kepone from adipose storage depots. The workers who were symptomatic were eventually treated with cholestyramine to interrupt enterohepatic circulation, representing the first successful use of the treatment.

Pentachlorophenol, colloquially called “penta,” is at the other extreme of structure, a very simple polychlorinated ring (see Figure 10.1). It is used as a wood preservative in pressurized applications where it keeps telephone poles, wharf pilings, and railroad tracks from rotting and disintegrating. Notwithstanding its appearance and its use outdoors, it is a non-persistent molecule that is degraded photolytically in sunlight, and by the metabolic action of bacteria and molds. Penta is only used industrially and is not available to individual consumers. Technical grade penta is contaminated with other chlorophenates of similar toxicity. Acute toxicity from penta results from uncoupling of oxidative metabolism and excessive heat production, resulting in hyperthermia. At lower exposure levels, it mimics thyroid hormone and is a suspected human carcinogen. In the past families have been exposed to it by burning railroad ties in a fireplace or through contamination of clothing.

### Pyrethroids

The pyrethroids are a chemical class of derivatives of pyrethrums, a class of natural plant toxins that are neurotoxic to insects. Pyrethrums are derived from chrysanthemums, and historically crushed chrysanthemum powder was used as a pesticide. In recent years, synthetic pyrethroids have been developed that have reduced

the price and vastly increased the use of the pyrethroids as a class. Pyrethrums were used almost exclusively for gardens and specialty crops. Now pyrethroids have emerged as a major class of commercial pesticides.

The inherent toxicity of the pyrethroids is minimal compared to other classes of pesticides, although the hydrocarbon solvents in which they are dissolved may also cause toxic effects. The pyrethroids are neurotoxic to insects and have low-level neurotoxic effects in humans, with symptoms of paresthesias (“skin crawling” is the popular description), cholinergic actions (manifested as salivation, vomiting, and fasciculation), and, rarely, seizures.

The most common problems associated with the pyrethroids are sensitization and allergic reactions, presenting as hay fever-like symptoms and accompanied by allergy to chrysanthemums. This reaction is more common with natural pyrethrum than with the synthetic pyrethroid derivatives. Treatment of allergic symptoms with antihistamines is usually sufficient. Severe toxicity or anaphylaxis is rare, and management is supportive.

## DEET

DEET is *N,N*-diethyl-3-methylbenzamide, a unique insect repellent applied topically at heavy concentrations. There is significant skin absorption. Although lipophilic, DEET is rapidly eliminated. It has been associated, at inappropriately high exposure levels, with neurological symptoms in children, such as ataxia and seizures. However, used as directed, DEET is thought to be acceptably safe. It became the center of controversy in the 1990s as a possible candidate for a toxic cause of the “Gulf War syndrome,” but the balance of evidence does not seem to support this.

## Phenoxyacetic Herbicides

The phenoxyacetic herbicides include several of the most heavily used pesticides, among them MCPA, dichlorprop, mecoprop, MCPB,

2,4-DB, and the virtually ubiquitous 2,4-dichlorophenoxyacetic acid (2,4-D), heavily used in agriculture and in household applications because it is effective and cheap, as it is no longer covered by patent. As a representative of the class and the compound of greatest current interest regarding toxicity, 2,4-D will be discussed in greatest detail.

2,4-D has been used heavily in agriculture, forest management, household lawn and garden care, and management of parks, grounds, and rights of way, including railway and highway sidings. 2,4-D is rapidly hydrolyzed in the environment, particularly under conditions of moisture and sunshine, disappearing to trace or undetectable concentrations within days. The compound was used as a defoliant during the Vietnam War in the form of "Agent Orange," in a mixture with the related herbicide 2,4,5-T (which has been withdrawn since then because of evidence of carcinogenicity in animal tests). Agent Orange and many of the earlier 2,4-D formulations were significantly contaminated with the dioxins in the manufacturing process; contemporary formulations contain little or no detectable dioxin.

2,4-D and the other phenoxyacetic herbicides are water-soluble compounds, dissolving poorly in oil or lipid. For this reason, they are not readily absorbed transdermally, especially in their acid form. Acute toxicity is therefore generally associated with ingestion due to an accident or suicidal intent. Chronic exposure results mostly from transdermal absorption during application in agriculture, however. Because 2,4-D is tightly bound to grass and leaves and degrades quickly in the environment, there is little opportunity for significant exposure after application is completed. Concern over the toxicity of 2,4-D centers on its potential for carcinogenesis.

The acute toxicity of the phenoxyherbicides to humans and other mammals is low. Rhabdomyolysis is an effect that appears to be unique to 2,4-D and can be quite severe, resulting in renal failure.

The carcinogenicity of the phenoxyacetic herbicides, particularly 2,4-D, has been a major issue in recent years. A large and somewhat contradictory body of evidence has emerged in the toxicologic and epidemiologic literatures; investigation of the question is complicated

by numerous and serious methodological obstacles. Taking into account the strengths and the weaknesses of the studies available, there is reasonably convincing evidence for a weak but causal association between exposure to 2,4-D in agriculture and the non-Hodgkin's lymphomas. Some evidence also suggests that this risk disappears when personal protection, such as coveralls and gloves, are used during spraying. The evidence is much less strong for an association between exposure and soft tissue sarcomas. There is no evidence and no suggestion that occasional household or lawn or garden use of 2,4-D carries such a risk, or that 2,4-D is associated with more common cancers.

### **Other Herbicides**

Among other agents, the herbicides are a particularly heterogeneous category of pesticide. Their toxicity in bulk use is a trade-off against traditional alternatives such as picloram (which was much more acutely toxic than any of the alternatives, especially glyphosate), and the risk of injury associated with hand clearing of weeds and unwanted vegetation, which is considerable.

#### ***Dipyridyl Herbicides***

Paraquat and the less potent diquat are herbicides that provoke cascades of free radicals intracellularly, resulting in lipid peroxidation. Human exposure to paraquat, by inhalation, skin absorption, or ingestion, causes severe lung toxicity leading to interstitial fibrosis which is usually fatal without lung transplantation.

#### ***Glyphosate***

Glyphosate is the most heavily used herbicide in recent years. It inhibits amino acid synthesis in plants, and is heavily used for postemergent (after the plant pokes out of the ground) broadleaf control. It has very low toxicity but it decomposes to formaldehyde,

which may cause some skin and mucosal irritation. Suicide attempts using glyphosate have been associated with multiple organ system failure.

## POLYCHLORINATED BIPHENYLS (PCBs)

The PCBs, as they are familiarly called, are a large class of compounds consisting of two six-membered aromatic carbon rings linked by a single carbon–carbon bond and substituted with chlorines (see Figure 10.1). All commercial preparations of PCBs were liquid mixtures of many congeners, the distribution of which may change with heating and aging of the mixture. PCB preparations are often contaminated with small amounts of dioxins and furans, discussed elsewhere in the chapter. The PCBs have very useful properties for industry, particularly as non-conducting cooling fluids for electrical transformers, laboratory and optical oils, plasticizers, and as vehicles for inks. In the past, occupational exposure was common and intense, with electrical workers often working with their arms immersed in the liquid and their clothes heavily contaminated.

The PCBs are very stable compounds and highly persistent in the environment, where they bio-concentrate in the food chain. For this reason, and from somewhat exaggerated concerns regarding human health, most uses of the PCBs have been phased out and new production is minimal. The extreme stability and persistence of the compounds means that they will be environmental pollutants for years to come and they are already among the most widely distributed synthetic compounds on earth.

Exposure to the PCBs in occupational settings is primarily transdermal, as the agents are readily absorbed by this route but are not very volatile. Being extremely lipophilic, they are accumulated, stored in adipose, and very slowly excreted; they appear in small concentrations in the breast milk of lactating women in many countries and urban areas.

There are 208 possible PCB compounds, defined by the number and placement of chlorines in the rings. For convenience, the PCBs

are referred to by a standard numbering system rather than by chemical name. Human toxicity depends on the composition and orientation of the PCBs in question. PCBs that assume a relatively fixed “flat” configuration, in two dimensions because of stearic hindrance by the way the chlorines are positioned, are called “coplanar.” They are more likely, in general, to show specific toxicity as endocrine mimics, to mimic the effect of thyroid hormone, and to show effects similar to TCDD (2,3,7,8-tetrachlorodibenzodioxin, the most potent dioxin) at much lower potency. PCBs (and dioxins) have been evaluated for toxicological similarity to TCDD and have been assigned toxic equivalency factors (TEFs, where TCDD = 1.0, a PCB might rarely be 0.1, and in the case of PCB-126, 0.0001, lower TEFs being more common). For purposes of risk assessment and toxicological characterization, mixtures of PCBs can be characterized by the sum of their mass weighted by the TEFs to yield a toxic equivalency (TEq). Other characteristics that predict a resemblance to TCDD are placement of chlorines in the *para*- position of the phenyl rings, two or more chlorines in a *meta*-position, and four or more total chlorines. Non-coplanar PCBs have a more heterogeneous toxicological profile.

The PCBs have been extensively studied toxicologically. Like the dioxins and furans, their effects appear to differ between humans (and other primates) and other animal species. They are, depending for potency on the mixture, promoters and possibly initiators of cancer in rodents and inducers of mixed function oxidase activity that results in abnormalities of porphyrin and steroid hormone activity. They are defined by the IARC as group 2A carcinogens, indicating limited evidence of carcinogenicity in humans. Human effects are known to include chloracne. (This is described in Chapter 14.) There is evidence suggesting that PCBs interfere with neurological development, presumably because of their thyroid-blocking effect. Infants born to mothers with relatively high PCB intakes from eating higher (but not excessively high) amounts of fish caught in the Great Lakes have had persistently diminished neurocognitive and scholastic performance as they mature.

PCBs are animal carcinogens in mixtures and potent tumor promoters. They are presumed to be human carcinogens, as a class. Unrecognized effects on humans cannot be ruled out and may remain to be discovered.

## POLYCYCLIC AROMATIC HYDROCARBONS

Polycyclic aromatic hydrocarbons (PAHs) are a family of multiple-ringed hydrocarbon compounds that are produced under conditions of incomplete combustion or of heat or pressure, and are therefore common in cigarette smoke, coal tar, smoke from fires, burnt food, particulate air pollution, and organic chemical processes. Several of them are known to be, or are highly likely to be, procarcinogens, presenting a cancer risk after absorption and metabolism. These hydrocarbon compounds include benzo(a)pyrene, chrysene, benzo(a)anthracene, benzo(b)fluoranthene, dibenzo(a,h)anthracene, dibenzopyrene, and 5-methylchrysene. They occur naturally in mixtures, never as individual compounds. Because of the difficulty in working with and chemically characterizing their complex mixtures, they are often monitored and quantified by the amount of one of their most common constituents, benzo(a)pyrene.

Structurally, PAHs are repetitions of a benzene (aromatic) ring in many configurations, with or without side groups. Naphthalene (two rings), phenanthrene (three rings staggered), and anthracene (three rings in a row) are the simplest members. Complexity increases as rings are added in straight chains (naphthacene with four rings, pentacene with five) or staggered (chrysene with four) or aggregated (pyrene with two in the middle and one on each side), and progress to rings of rings (coronene, ovalene). The chemical reactivity depends on how many rings are next to one another sharing electrons, and how many are at the extremities pulling electrons away, to form a resonant aromatic structure. The less aromatic the ring within the structure, the more reactive it will be for reactions such as Phase 2 metabolism. The “aromaticity” (electron density) of the ring structure also determines the optimal frequency of ultraviolet light absorption.

Photoabsorption is an important characteristic of most PAHs and many fluoresce.

As a class, PAHs are easily inhaled and ingested; they are relatively efficiently absorbed through the skin but also have local effects on skin. PAHs are distributed by the circulation and metabolized by the mixed function oxidase system for Phase I and II metabolism, then are mostly excreted in urine and feces.

The metabolites of PAHs are highly active biologically. They are important carcinogens, as noted, affecting skin, lung, bladder, and kidney, and act as both mutagenic initiators (interacting directly with DNA) and potent promoters. They bind to the Ah nuclear receptor in cells, but at a lower affinity than dioxins and PCBs. They are potent inducers of the mixed function oxidase in all tissues, especially lung and liver, and therefore may accelerate the Phase I metabolism and activation of other carcinogens.

Because of their ultraviolet absorption characteristics, they are photosensitizing, increasing skin damage in sunlight, and causing dermatitis among exposed workers such as roofers.

PAHs are also highly toxic to aquatic species. Some have endocrine disrupting effects.

## SILICA

Silica is silicon dioxide ( $\text{SiO}_2$ ), the material of quartz and glass in different states. In the form of quartz, it is one of the most widespread minerals on the surface of the earth. Silica exists in several forms, or “polymorphs” mineralogically, of which the most important division toxicologically is between amorphous silica in forms such as diatomaceous earth, which is relatively non-fibrogenic and extensively used as a filtering material, and the much more fibrogenic crystalline forms of silica. These more fibrogenic crystalline forms of silica take the form of quartz (technically,  $\alpha$ -quartz) and “calcined” (heat-treated) silica (cristobalite or tridymite, both of which are more potently fibrogenic than  $\alpha$ -quartz). Glass has different properties altogether and is discussed as a minor topic in the section on asbestos in this chapter.

$\alpha$ -Quartz, by far the most common form of crystalline silica, is ubiquitous in the environment. In the form of large particles, it cannot be inhaled into the deep lung and is therefore rarely an environmental hazard. It becomes an occupational hazard when dust is generated during operations where rock is drilled, cut, crushed, or mined, where high-temperature-resistant refractory brick is used, or where silica flour (finely divided powder) is used as an abrasive or filler material. Silica dust is most potently fibrogenic when it is freshly cut or crushed (exposing fresh reactive groups on the surface of the dust particles).

Because of its importance as a hazard in mining, silica dust is one of the oldest recognized occupational hazards. Of all occupational lung diseases, silicosis is among the most common and most dreaded. Known since ancient times, silicosis constantly re-emerges as a health hazard as technology changes, most recently in construction or caisson work, tunneling, ceramic products, glassmaking, silica flour, and sandblasting, as well as in the traditional occupations of mining, grinding, and stone polishing.

Crystalline silica dust, inhaled into the alveoli and engulfed by macrophages as described in Chapter 2, is an intensely fibrogenic dust. The determinants of silica's toxicity *in vitro* include the concentration and duration of dust exposure, particle size distribution, the crystalline isomorph, and the presence of a freshly fractured surface (which exposes charged groups and increases cytotoxicity). It is the cause of the pneumoconiosis known as silicosis (discussed more extensively in Chapter 17) and has been associated with bronchogenic carcinoma. Silica is classified as a Class I human carcinogen by the IARC. The pathophysiology of silicosis includes the penetration of the silica particle to the lower respiratory tract, phagocytosis by macrophage and subsequent cytotoxicity, cytokine elaboration, concurrent B and T cell activation and proliferation, a concomitant inflammatory response, and finally, fibroblast activation and fibrosis. Cytokine release, free radical formation, and intracellular DNA damage are associated with carcinogenicity, probably by a different mechanism from that associated with asbestos.

Silicosis is an especially pernicious pneumoconiosis that initially does not impair pulmonary function but, in its most advanced stages or when complicated by tuberculosis, behaves in some respects like a non-metastasizing cancer and may carry a similarly bleak prognosis. Silicosis is discussed in greater detail in Chapter 17. The most famous incident involving an epidemic of silicosis was the “Hawk’s Nest” disaster of the early 1930s, an incident involving dry boring of a railroad tunnel in Gauley Bridge, West Virginia, in the absence of even rudimentary dust control measures for the workers. (Managers and engineers were provided with dust masks.) An estimated 1,500 impoverished workers, mostly rural African American men, died of accelerated or acute silicosis, or were permanently disabled with chronic forms of silicosis in this Depression-era tragedy.

Within the past ten years, silica has also been confirmed by IARC to be a Class I human carcinogen. Silica-associated lung cancer has been shown to be associated with fibrosis in epidemiologic studies; also, silicosis appears to increase cancer risk. An increased incidence of silica-associated lung cancer is associated with the profusion of small opacities, heavy exposure to silica dust, and tobacco exposure. Confounding of effects occurs when the patient is a cigarette smoker, or has been exposed to arsenic or radon and its daughters.

Other silica-related diseases include Erasmus syndrome (a form of systemic sclerosis that may be found in silicotic patients), other auto-immune disorders, chronic nephritis, silica nephropathy (with glomerulonephritis resembling lupus nephropathy), and other airway diseases, as the silicotic inflammation results in airway obstruction.

Control of exposure to silica is mostly a matter of dust suppression. Technically, silica exposure is easily prevented by relatively simple means, such as wet cutting of stone and substitution of sand in sandblasting by one of the many inexpensive alternatives. In those industries where the problem has been long recognized, such as mining, substantial progress has been made in controlling exposure.

## SOLVENTS AND VAPORS OF VOLATILE HYDROCARBONS

For purposes of occupational and environmental toxicology, and as used in industry, the working definition of a solvent is a liquid organic compound that dissolves other organic compounds. It is lipophilic and usually volatile. This is not the same definition that would be used by a chemist. Ironically, the most common and important solvent of all, water, is excluded from this working definition.

Many of the same compounds that are solvents in bulk are encountered in other contexts in OEM. For example, the trihalomethanes are common solvents, but they are also important disinfection by-products generated in drinking water at low levels by the interaction of free chlorine and trace organic matter. TNT, much better known as an explosive, is also a solvent. Other solvents may be used as fuels, feedstocks, drugs of abuse, beverages, or antifreeze.

Solvents are ubiquitous and critically important in industry. They are used to dissolve materials into a liquid phase, to clean parts, to extract oils and greases, to remove grease, for dilution and dispersal, for dry cleaning, and as carriers for inks, dyes, paints, and coatings.

Solvents belong to many chemical classes, as outlined in Table 10.9. The most common solvents are short-chain aliphatic (straight-chain) hydrocarbons, usually substituted, and low-molecular-weight aromatic (ring structure) hydrocarbons. Alcohols and glycols are somewhat less common, although also heavily used. These are discussed elsewhere in this chapter.

Although a highly heterogeneous category of agents, solvents have several common physicochemical characteristics and toxicological properties. In addition, specific solvents have their own intrinsic properties as well. The remainder of this discussion will emphasize those solvents that have a relatively high vapor pressure at room temperature and that predominantly partition to oil or lipid rather than water. These are the majority of solvents in common use, excluding water.

Volatile solvents are absorbed rapidly by either the inhalation or the transdermal route in occupational exposures. By the inhalation route,

**Table 10.9.** Chemical Classes of Solvents

Class	Example(s)
Aliphatic hydrocarbons	Gasoline, hexane
Cyclic hydrocarbons	Cyclohexane
Aromatic hydrocarbons	Benzene, toluene
Ketones	Cyclohexanone, acetone
Aldehydes	Acetaldehyde
Alcohols	Ethanol, methanol
Ethers	Glycol ethers
Esters	Ethyl acetate
Nitrohydrocarbons	Ethylnitrate, trinitrotoluene (TNT)
Halogenated alkanes	1,1,1-trichlorethane, tetrachloromethane (carbon tetrachloride), trichloromethane (chloroform)
Halogenated alkenes	Trichloroethylene, vinyl chloride
Mixtures	Kerosene, “white spirit,” “petroleum distillates”

solvent vapors diffuse across the alveolar-capillary barrier, enter blood, and are carried to the periphery, where they may preferentially partition to tissue containing a high lipid content, such as brain, liver, and myelinated peripheral nerve tissue. Transdermal absorption is significant and rapid for many solvents, which pass readily across the lipid-containing dermis. They may also damage the skin by de-fatting, further speeding absorption.

Solvents are typically stored in depots of high lipid content, mostly adipose tissue. Depending on the partition, and on the coefficient properties between lipid and aqueous blood and between blood and air, the solvent will tend to leave its tissue stores and be carried by the bloodstream back to the lungs. It then passes back across the alveolar-capillary membrane to be exhaled, at a greater or lesser rate. For some solvents, such as toluene, the rate is rapid, and measurement

of end-expired breath toluene is an excellent indicator of tissue stores. For diethyl ether, however, the process is very slow. Bio-transformation is usually by the mixed function oxidase-cytochrome system, through various reactions. This results in activation to toxic metabolites for some, especially halogenated hydrocarbons. Expiration, unchanged, is a major route of excretion for many solvents with rapid kinetics of redistribution. Conjugated products are typically excreted in urine and feces. Biomonitoring can usually be undertaken by measuring post-shift in either expired air or urine.

The toxicodynamics of solvents reflect their lipid solubility in the cell membrane and their propensity to form free radicals after undergoing bio-transformation. As a rule, these effects tend to be strongest for brominated analogues of solvents (seldom used in industry), strong for chlorinated solvents, and stronger with the degree of substitution, and less for unsubstituted solvents. They are variable in their cancer risk.

The common toxicological action of these solvents is narcosis. At high levels of exposure, solvents act like general anesthetics, depressing central nervous system functions and manifesting their effect through behavioral changes (euphoria, lightheadedness, disorientation, mood changes, slurred speech, and confusion), through agitation and disinhibition (leading to impulsive behavior), dyscoordination and ataxia, and progressing in extreme cases to loss of consciousness, respiratory depression, seizures, and cardiac arrest. The syndrome is familiar from ethanol intoxication, of course. The mechanisms of solvent-induced narcosis is physicochemical, probably related to dissolution in the lipid bilayer of the membranes of nerve cells, and is the same mechanism as that of general anesthetics. Indeed, some solvents were once used as general anesthetics, such as diethyl ether and chloroform. Fatal solvent exposures usually occur in the context of a confined space with failure of personal protection or, more rarely, substance abuse.

A second toxicological property of solvents is chronic neurotoxicity. Several solvents have the potential to induce central neurotoxicity at levels encountered in the workplace, with chronic effects on cognition

(usually manifested as reduced short-term memory, mental confusion, and disorientation), visual changes (acquired color-blindness, reduced hearing acuity), mood changes (depression, irritability, malaise), and motor changes (relatively coarse tremors of the hands, reduced coordination, reduction of strength particularly in the hands). The solvents most commonly associated with these changes include toluene, trichloroethylene, gasoline (a mixture of aliphatics and aromatics), and carbon disulfide (rarely used as a solvent today because of its extreme toxicity), but other solvents are likely to be implicated, judging from the rapidity with which the literature on solvent-induced central neuropathy is expanding. It is possible that this is a general effect, present to some degree in all organic solvents. The behavioral, affective, and cognitive changes constitute the so-called "painter's syndrome" because the constellation of findings is observed often among workers in that occupation, particularly spray painters. It is difficult to document objectively in individuals, although changes on neuropsychological testing can be demonstrated in groups. Special batteries of test have been created to evaluate chronic neurotoxicity in occupational groups, particularly the World Health Organization (WHO) Neurobehavioral Test Battery. A major drawback in studies of these effects is that exposures to solvents in most occupational settings are multiple and intermittent, and it is difficult to identify and separate the effects of a specific solvent. It is also difficult to adjust for ethanol intake.

Peripheral neurotoxicity is an effect of organic solvents, although the mechanisms appear to differ and the potency varies widely. The most potent known peripheral neurotoxins have been removed from the workplace as solvents, including carbon disulfide and those compounds metabolized to a potent neurotoxin highly toxic to axons: 2,5-hexane dione, *n*-hexane, *n*-hexanone, methyl-*n*-butyl ketone, and tri-ortho-cresyl phosphate. Methyl-ethyl-ketone (MEK) is a commonly used solvent that is itself rather innocuous, but potentiates the toxicity of MnBK and other 2,5-hexanedione precursors. Styrene and tetrachlorothethylene are suspected of neurotoxicity. The presence of lead in leaded gasoline used as a solvent poses a

considerable risk for neurotoxicity over and above any contribution from the hydrocarbon components. It is thought by many investigators that more solvents are associated with peripheral neuropathies than is commonly recognized.

As a rule, solvents also share hepatotoxicity and nephrotoxicity, to varying degrees.

Hepatotoxicity due to solvents is generally a result of dehalogenation and free radical formation, which is why more highly chlorinated hydrocarbons carry higher toxicity. Tetra- and trichloromethane are most potent and have been used experimentally to study mechanisms of liver damage. The hepatic effect is stereotyped, involving hepatocellular injury with release of transaminases, steatosis (fatty liver), and occasionally progressing to hepatic necrosis or cirrhosis in extreme cases. This effect may be severe with heavy exposure, leading to fatty liver changes, hepatic necrosis, and fibrosis. The mechanisms appear to be similar in each, and some of these same compounds (chloroform, carbon tetrachloride, vinyl chloride, dichloromethane, and vinylidene chloride) are known or suspected hepatic carcinogens. The solvents most commonly implicated in severe hepatotoxicity include chloroform, carbon tetrachloride, 1,1,2,2-tetrachloroethane, 1,2-dichloroethane (ethylene dichloride), 1,1,2-trichloroethylene, and, of importance because of the frequency of its current use rather than its intrinsic toxicity, 1,1,1-trichloroethane. The counterpart compounds substituting bromine or iodine for chlorine have, in general, similar hepatotoxic properties. While solvents are in high concentration in the circulation, they tend to competitively inhibit the metabolism of other xenobiotics that share common pathways, such as acetaminophen, conferring serious risk with mixed exposures.

Solvent-induced nephropathy may reflect a more diverse array of pathological states. Acute tubular necrosis can occur after massive exposure to halogenated hydrocarbons, toluene, and petroleum distillates. Glycol ethers may cause acute renal failure by a different mechanism, as well as oxalate formation and tubular obstruction. Glomerulonephritis may or may not occur with chronic long-term exposure to gasoline; this is debated.

Organic solvents are also de-fattig agents that remove essential oils and lipids from the skin. With prolonged or repeated exposure, they all are capable of causing a dermatitis that can become quite severe, with fissures and cracking and, with repeated exposure, hypertrophy of the dermis resulting in a irreversible “nummular dermatitis.”

Halogenated hydrocarbons, including some chlorofluorocarbons, are usually potent cardiac sensitizing agents. Prolonged and intense exposure may lead to arrhythmias and cardiac arrest.

Individual solvents also have their own toxicity characteristics that are important clinically. These are summarized in the remainder of this section, in alphabetical order.

### **Benzene ( $C_6H_6$ )**

Exposure to benzene, extensively used in the past as a solvent, today is highly reduced because of highly restrictive occupational standards. It is present in relatively high concentrations, about 6 percent, in gasoline, particularly in unleaded gasoline. It is highly volatile, with a simple aromatic ring structure, and probably shares many of the common toxicological properties of the organic solvents, including neurotoxicity, but is characteristically and predominantly a toxin to bone marrow. Benzene appears to be unique among the low-molecular-weight aromatics in inducing blood dyscrasias. In marrow, an activated metabolite of benzene causes maturation arrest. Some elements escape as malignant clones, resulting in acute myelogenous leukemia (AML), the signature cancer associated with benzene. Whether benzene can induce other leukemias is controversial. All blood cell-forming elements are depressed, but initially granulocytopenia, lymphocytopenia, thrombocytopenia, and anemia may appear individually or in progression, in more or less that order. Pancytopenia results from prolonged exposure and aplastic anemia, and in some cases, probably myelofibrosis. When necrosis of marrow occurs, the aplastic anemia is usually irreversible, although occasionally withdrawal from exposure may lead to recovery if marrow suppression is not too far advanced. Subsequent re-exposure is

thought to carry a high risk of irreversible aplastic anemia after initial recovery.

Benzene is monitored by urinary benzene or phenol and is tightly regulated. Because of its carcinogenic potential and risk, benzene almost always dominates a risk assessment in which it is present as a hazard. For example, benzene, toluene, ethylbenzene, and xylene, which collectively are called “BTEX,” are commonly found together at contaminated sites. Risk assessment conducted to determine whether the level of remediation is adequate is almost always driven by the benzene component.

### **1-Bromopropane**

1-Bromopropane (1-BP) was introduced to replace chloroethane solvents and perchlorethylene, which deplete the stratospheric ozone layer. 1-BP has much less environmental impact because its half-life in the atmosphere is short. It was rapidly adopted for use as a general solvent and for degreasing, as well as in dry adhesives, foam cushion manufacturing, and dry cleaning. 1-BP is now widely used despite the absence of formal OSHA standards or NIOSH recommended exposure guidelines. The American Conference of Governmental Industrial Hygienists (ACGIH) has recommended a TLV of 25 ppm as an 8-hour time-weighted average. Recently cases of 1-BP exposure at relatively high levels of exposure have been reported, characterized by severe and painful peripheral neuropathy and joint pain, associated with spastic paraparesis, and preceded by acute illness characterized by confusion, nausea, and headache. 1-BP is associated with severe axonal degeneration. In at least one case, there appeared to be chronic neurocognitive changes. Levels of exposure associated with this level of toxicity are thought to be around 130 ppm. 1-BP is also irritating to mucous membranes.

### **Chloroethanes**

Chloroethanes are widely used in many industries as solvents, degreasing agents, and fumigants. Four chloroethanes have been

implicated as potential human carcinogens on the basis of animal studies: 1,2-dichloroethane (ethylene dichloride), 1,1,2-trichloroethane, 1,1,2,2-tetrachloroethane, and hexachloroethane. NIOSH recommends that exposure to these and related compounds be kept to a minimum. Additionally, 1,2-dichloroethane, which is not only a solvent but a chemical waste material from vinyl chloride manufacturing, is associated in animal studies with hypotension, pulmonary edema, hepatotoxicity, nephrotoxicity, and adrenal cortical necrosis. It has been used primarily as a cleaning and degreasing agent and as a fumigant. Other chloroethanes have similarly diverse spectra of toxic effects. Oxidation products of the chloroethanes include hydrochloric acid and phosgene, a highly toxic agent that may cause acute nephrotoxicity (see Chapter 17).

### **Chloromethane and Other Halomethanes**

Chloromethane (methyl chloride,  $\text{CH}_3\text{Cl}$ ) and its brominated and iodinated counterpart halomethanes are used in fire extinguishers and as refrigerants, propellants for aerosols, blowing agents for plastic foam, fumigants, and degreasing agents. Their chemistry has led to serious concern over environmental effects from depletion of the stratospheric ozone layer by reactions with halomethanes released into the atmosphere. Their use is being phased out under the terms of an international convention. Occupational exposure should also be kept to a minimum due to possible cancer and reproductive risks. All of these halomethanes have been found to induce cancer in animals; chloromethane has also been shown to induce birth defects in the fetus of pregnant mice exposed to the agent.

### **Dichloromethane**

A widely used ingredient in paint strippers (along with methanol), dichloromethane (methylene chloride,  $\text{C}_2\text{H}_2\text{Cl}_2$ ) is metabolized in the liver to carbon monoxide, which is discussed elsewhere in this section. The agent has also demonstrated carcinogenicity in animal studies.

### Tetrachloroethylene

Tetrachloroethylene (perchloroethylene or “perc,”  $\text{CCl}_2=\text{CCl}_2$ ) is extensively used as a dry cleaning agent; less commonly it has been used for fumigation and degreasing. It has very desirable properties and has replaced other solvents that are more flammable and acutely toxic; its volatility is relatively low. It is capable of producing pulmonary edema at high exposure levels. It may also induce neurotoxicity and chloracne with sufficient exposure (see Chapter 17). Perc is also an animal carcinogen. In general, exposure in the workplace is limited, including dry cleaning establishments, because it is expensive and is extensively recovered and recycled to minimize loss. Uncontrolled exposure to perc resulting in toxicity is very unusual. Exposure in excess of occupational exposure standards usually occurs as a result of poor maintenance and inadequate ventilation in dry cleaning establishments. Toxicity is more often related to direct handling of the agent. Skin absorption may be a more important potential route of exposure for this agent. Perc has been a significant agent of groundwater contamination from spills, but because it can be recycled and has commercial value in dry cleaning, little goes to waste. Evaporated perc is a significant contributor to ambient air pollution, participating as a catalyst in many photochemical reactions. Perc also acts as a catalyst to deplete ozone in the stratosphere, but most perc does not reach that altitude because of atmospheric reactions at lower levels. Recently, several states, including California and New Jersey, have proposed banning perc, to prevent its emission as an air toxic.

### Toluene

Toluene (methylbenzene) is a lipid-soluble aromatic hydrocarbon less volatile than benzene, with which it is sometimes contaminated. It is heavily used as a solvent for paints, lacquers, and glue, and as thinner for paints and coatings. It is also used as a feedstock for chemical production. It is one of the most potently neurotoxic solvents; repeated heavy exposure is associated with chronic neurological sequelae, including memory loss, ataxia, labile affect, and tremor. Toluene is occasionally abused by

those who engage in “glue sniffing” (usually in the form of “huffing,” repeated inhalation and re-breathing into a plastic bag), with high levels absorbed in a short time and resulting chronic central neurotoxicity after repeated use. The clinical picture of toluene toxicity has developed principally through studies of such abuse and by its contribution to “painter’s syndrome” in spray painters. Toluene itself does not appear to be associated with peripheral neuropathic changes. Other known effects include dermatitis, caused by the de-fattening effect, and cardiac arrhythmia resulting from myocardial sensitization. Urinary hippuric acid excretion may reflect toluene (or styrene) exposure, but this is highly variable in normal subjects due to dietary sources and endogenous metabolism.

### **Trichloroethylene**

Previously widely used as a degreasing agent, trichloroethylene (“TCE,”  $\text{CH}_2\text{ClCCl}_3$ ) has virtually disappeared in North America because of its suspected carcinogenicity, based on animal studies, and other effects. These include neurotoxicity and a peculiar vasodilatory effect resulting in facial and truncal blushing popularly called “degreaser’s flush,” a phenomenon made worse by consumption of alcoholic beverages, similar to the interaction between ethanol and disulfiram. It has been replaced by 1,1,1-trichloroethane.

### **Xylenes**

Xylenes are a family of three dimethyl-benzene rings (ortho, para, and meta isomers) without halogen substitutions. Xylene is exceptionally irritating to mucous membranes. It frequently causes eye irritation and nasal discomfort, and sometimes cough and bronchial irritation when used in poorly ventilated areas.

## **TRACE ELEMENTS**

“Trace elements” are metals and metalloids (selenium and arsenic) that are normally present in very low concentrations in the body. Iron and magnesium are not trace elements because

substantial quantities are normally present in the body. Many such metals are essential to various metabolic functions (chromium, cobalt, copper, manganese, molybdenum, selenium, vanadium, zinc), but only in minute quantities. Deficiency conditions of these trace elements are rare because they are needed in such small amounts.

In occupational medicine, trace metal analysis may be used to monitor exposure or to evaluate suspected toxicity. In environmental medicine, it may be used to track body burden in populations or to evaluate particular patients, usually children, for lead toxicity.

Trace metal analysis of body fluids reflects the body burden, or internal dose, coming from all sources. There is no way to separate out, or apportion, sources of exposure. There are many sources of exposure for these trace elements in daily life, particularly in foods, but only in extremely low amounts. Occupational exposures by inhalation or ingestion are usually significant because they typically occur in much higher amounts than dietary intake and therefore are detectable as a marked, temporary increase in intake. Supplemental vitamins (containing chromium, for example), a diet rich in fish (which may contain organic forms of arsenic that are measurable but not toxicologically significant, or of mercury), and cigarette smoking (cadmium) are common sources of unusually high but not toxic levels of trace elements.

Several trace elements, most important arsenic and nickel, exist in different forms, or “species,” and the use of chemical tests to separate them is called “speciation.” This is important for arsenic especially, because high urinary arsenic levels may reflect arsenobetaines, organic arsenic compounds derived from fish and shellfish, which have very low toxicity and do not cause arsenic toxicity. Speciation isolates inorganic arsenic, if there is a question about arsenic intake that for some reason cannot be answered by a dietary history. Inorganic arsenic is associated with a risk of lung, skin, and bladder cancer; it is also associated in cases of long-term exposure at high levels with a characteristic skin rash and hyperkeratosis, and has protean symptoms in acute toxicity. Nickel metal is a potent allergen but not

highly toxic; however, the species nickel subsulfide, which is mainly found in smelting, is a potent carcinogen, and nickel carbonyl is acutely toxic. Chromium VI, the hexavalent species, is carcinogenic, but chromium III, the trivalent species, is not.

The metals of concern vary greatly in their toxicokinetic behavior. Manganese is efficiently absorbed only by inhalation and also accumulates in red cells (96 percent), but it is predominantly excreted in bile and feces and has a very short half-life (only days). There is a very close correlation between intake of lead and its concentration in body fluids, while there is a very poor correlation for manganese. Thus, it is necessary to know as much as possible about the behavior of the specific trace element in question before attempting to interpret the toxicological significance of an elevated trace metal profile analysis.

The principal challenge in interpreting levels of trace elements is that there is a wide gap for most between the highest levels commonly seen in the reference population, or even exposed workers, and the toxic threshold, which is often very poorly characterized. The major exceptions to this are lead, mercury, and manganese. In each of these three cases, exposure as reflected in normal, unexposed community residents is within one or two orders of magnitude below toxic levels for the metal.

The clinical interpretation of trace element analysis has lagged behind the technology, and these tests are often ordered blindly, with no advance plan for clinical intervention. Interpretation requires some knowledge of the toxicokinetics of each metal of interest and the preferred medium for analysis for each: serum, whole blood, or urine (24-hour collection only). Clinical interpretation is often difficult because even marked elevations measured against an arbitrary reference range may lack clinical or toxicological significance.

For these tests, the reference ranges typically reflect low levels of exposure in the general population and do not apply to metal workers or to normal individuals in the population with special habits or vocational exposure. It is to be expected that workers handling metals in occupations such as welding and industries such as steelmaking

will have higher levels than will an unexposed member of the general population. Trends are often more informative than are concentrations at one time.

As a practical matter, the greatest interpretation problems tend to be found with manganese because serum levels have a poor correlation with both recent exposure and neurological symptoms. Molybdenum and vanadium are often found to be elevated among workers exposed to metals who show no evidence of clinical illness. (The toxicological significance of vanadium lies mostly in its irritant effect on bronchial epithelium.) Interpretation of the trace element profile analysis overall, when an elevation occurs, generally requires close attention to the pattern of elevation, clinical context, absolute and relative magnitudes of the elevation, and knowledge of the exposure history.

The routine use of biological monitoring to evaluate exposure to workplace hazards is the basis of “biological exposure indices” (BEIs), such as those developed by the ACGIH. (See Chapter 7.) An absolute requirement for an acceptable BEI is that it correlate closely with documented exposure in the workplace for individuals. This is true for many metals, such as lead and cadmium, but not for others, such as manganese. Manganese levels do reflect workplace exposure on an average, group basis, but for any one individual the levels can be highly variable despite constant exposure.

The standard technology for trace element analysis is called inductively coupled plasma mass spectroscopy (ICP-MS). It is an exceptionally accurate and specific method for low concentrations of trace elements, in the nmol/l or  $\mu\text{mol/l}$  range. Three biological fluids are assayed on a routine basis: serum, whole blood, and urine. Hair measurement is unreliable and is rarely used in toxicology except for historical and forensic research. Hair is easily contaminated.

The final component of a trace element analysis is a profile of twelve to fourteen trace element concentrations in any or all of the three body fluids, expressed in one of two ways. In the United States, concentrations are expressed in  $\mu\text{g/l}$ ,  $\text{ng/l}$ , or, for lead only,  $\mu\text{g/dl}$ , in accordance with the OSHA lead standard. In most of the world, trace element analysis is reported in SI units of micromoles per liter for

those in relatively higher concentrations, and in nanomoles per liter for those in lower concentrations.

Clinical testing is performed on serum for metals that are carried in the blood in dissolved form or that are bound to serum proteins: aluminum, antimony, barium, beryllium, copper, manganese, nickel, selenium, vanadium, and zinc.

Clinical testing is performed on whole blood for metals that, while present in serum, are mainly concentrated in the red cell fraction of blood. Those metals that accumulate preferentially in red cells are not accurately reflected in serum concentrations, and only whole blood concentrations are valid for these metals: cadmium, cobalt, molybdenum, lead, and thallium. Copper, for example, will be underestimated in whole blood, but copper toxicity requires a defect in storage and transport and is not of practical importance except in Wilson's disease, which is diagnosed by a different test.

Urine is tested to determine the excretion of metals, ideally over a 24-hour period. This is usually the most accurate reflection of the total body burden of the metal. Many factors affect excretion of metals over short periods: state of hydration, renal function, intake with foods, short-term exposures from other sources, and renal blood flow. Over a longer period of time, however, these variations even out, and excretion is then generally directly related to the average serum concentration during this period (in equilibrium with red cell concentration, in the case of those metals that accumulate in red cells). Urinary excretion over 24 hours is used to determine concentrations of these metals: aluminum, antimony, arsenic, barium, beryllium, bismuth, cadmium, copper, manganese, selenium, lead, thallium, vanadium, and zinc. Spot urines are not as helpful because concentrations vary widely, both those of the trace elements in urine and the concentration of the urine itself. Normalization for specific gravity or urine concentration of creatinine has been used to correct for concentration as a substitute for 24-hour collection, but it is a poor substitute.

A 24-hour urine collection program is not practical in the workplace, however. Therefore, the BEIs established by the ACGIH are based on random spot urine samples, often collected at a specific time in relation

to exposure (pre- or post-shift). To compensate for variable urine output, the measured concentration of the trace metal is divided by the concentration of creatinine in the sample. Thus, the BEI for cadmium in urine using a creatinine correction is 5 µg/g creatinine.

The results from trace element analyses on occupationally exposed populations can be compared with the reference range only as a very rough guide. A published reference range is usually used by default. The reference range for trace elements means something very different from the “normal range” of a clinical test as it applies to a specific population group.

Trace elements, as with most environmental exposures, are characterized by a log-normal distribution, not a normal distribution, as is the case for the more common clinical laboratory tests. The normal range of a clinical test is highly meaningful; it represents the range of a biochemical parameter representative of healthy people, and is normally set to include 95 percent of the distribution for subjects in presumptive good health, without indication of a clinical disorder. A finding outside this range suggests a deviation from the normal homeostatic mechanisms of the body that keep the internal chemistry within the bounds of optimal function. The reference range for a trace element analysis, on the other hand, has no such meaning. It indicates only the range of levels of a trace element that would be expected from the usual exposure sources encountered in daily life, with no allowance for occupational or other special exposures. The terms “high,” “medium,” and “low,” therefore, are not meaningful, and when found on reports suggest that the laboratory is unfamiliar with trace element toxicology. Such descriptors certainly do not imply a need for chelation, which is a common focus of inquiry by physicians who receive the reports and do not know what to do with the information.

Toxicity thresholds have not been established for most trace elements and probably never will be. Thresholds may vary according to the age and genetic susceptibility of the individual, on whether the accumulation took place over sufficient time for tolerance to develop, on whether it was from food intake, on whether other exposures have occurred, on whether there is a coexisting clinical disorder, and

other factors. Published toxicity levels are approximate and need to be cautiously interpreted in context.

One useful clue to interpreting trace element analyses is the pattern of elevations in the profile. Certain elevations clearly go together: molybdenum, manganese, vanadium, and selenium, for example, are often elevated or relatively elevated together in welders compared to the reference range. There is no obvious toxicological significance to this observation. Rather, it probably reflects the combined exposure encountered in certain industrial processes.

Certain elevations above the reference range have well-described associations that have no toxicological significance. Arsenic, for example, is often elevated following a meal of seafood because shellfish accumulate organic arsenic compounds of low toxicity. Nickel and cadmium may be elevated among cigarette smokers. Selenium, chromium, and zinc may be elevated when the subject is taking supplements, usually from health food stores. Aluminum is routinely observed to be elevated in individuals in renal failure on dialysis, and correlates with years on dialysis. Substantial elevations of thallium are very unusual and might suggest ingestion of rat poison.

Perhaps surprisingly, trace element analysis is often unreliable in acute exposures. Metal fume fever, which is caused by inhaled zinc, copper, and possibly cadmium (not to be confused with cadmium pneumonitis, a distinct and life-threatening condition), seems to be a cytokine response to a transient bolus of metal that does not persist and cannot be confirmed by trace element analysis after symptoms begin. On the other hand, the acute, usually lethal forms of pneumonitis that are associated with inhaled cadmium and mercury fumes do seem to be associated with elevated levels. Acute toxicity by arsine, which induces hemolysis and is usually fatal, is not diagnosed by arsenic levels.

Trace element profile analysis is an imperfect tool for clinical screening. It is a cardinal principle of medicine that tests should not be ordered unless the requesting physician knows how to interpret the results. Unfortunately, trace element analysis is very often used as

a screening tool without a clear picture of its appropriate application, particularly in nutritional and alternative medicine.

Many questionable laboratories offer trace element analysis, some by mail order, and provide reports of dubious merit. Such laboratories require further investigation for quality assurance before being entrusted with samples. Features that might identify such a laboratory include testimonials on the Web site, advertisement for nutritional supplements, offers to provide medical counseling by e-mail, or to provide value judgments (such as "high" or "toxic"). Such a questionable laboratory might also list symptoms associated with toxicity of various elements, omit specification of detection limits, and fail to distinguish between the upper reference range and the toxicity threshold. Very often these reports are color coded and report trace element levels as a percentile of the population, which is usually meaningless.

On the other hand, conventional laboratories in mainstream medicine often perform trace element analyses individually, charging for each one. This makes no sense analytically because ICP-MS is used anyway and could give the entire profile with no additional effort. Most commercial laboratories simply do not want to be bothered with a more responsive service because trace element analysis is a low-profit business, with many expenses related to quality assurance for the small volume served. They may also be eager to avoid the complications and potential liability of dealing with the unconventional practitioners that this line of business attracts. This reluctance to provide the service does complicate the practice of toxicology in OEM, however.

Trace element analysis is best used for certain specific applications such as establishing exposure and BEIs, and to confirm an association following a compatible diagnosis. As experience with the method grows, trace element profile analysis will become increasingly useful. For the moment, however, it should be used cautiously and the OEM physician should order the test only with a clear idea of why he or she is doing so, and what he or she will do with the results.

## **RESOURCES**

*A contemporary reference book or handbook of toxicology should be on the shelf of every OEM physician. There are many titles available, and the physician should choose the format he or she is most likely to use and with coverage most appropriate to the specific practice.*

Hathaway GJ, Proctor N. *Proctor and Hughes' Chemical Hazards of the Workplace*. 5th ed. Philadelphia: Wiley-Interscience; 2004.

National Institute for Occupational Safety and Health. *NIOSH Pocket Guide to Chemical Hazard*. Atlanta: NIOSH, Centers for Disease Control. Available online at <http://www.cdc.gov/niosh/npg/>. [Handy guide to toxicity and standards. Updated periodically. Use current edition.]

Shannon MW, Borron SW, Burns M. *Haddad and Winchester's Clinical Management of Poisoning and Drug Overdose*. 4th ed. Philadelphia: Saunders; 2007.

Sullivan JB, Krieger GR. *Clinical Environmental Health and Toxic Exposures*. 2nd ed. Philadelphia: Lippincott Williams and Wilkins; 2001.

## **NOTEWORTHY READINGS**

Agency for Toxic Substances and Disease Registry Toxicological Profiles. Available online at <http://www.atsdr.cdc.gov/toxpro2.html>. [A series of publications on selected chemicals. The Profiles are the most comprehensive and accessible sources of information.]

# 11 BUILDINGS

Building-associated outbreaks are among the most common workplace problems that occupational and environmental medicine (OEM) practitioners face. The setting for these incidents is usually office buildings, often schools, and occasionally homes. They occur in buildings with no unusual industrial activity, chemical hazards, or significant emissions sources other than kitchens. The buildings are usually more recently built or renovated, and they are sometimes quite new.

These outbreaks are often puzzling. The workers who are affected present subjective symptoms, often having a history of sinus or allergy problems, and they are usually passionate in their response to the perceived health threat. The OEM physician is called in because of the symptoms, on the assumption that this is an outbreak of disease associated with poor indoor air quality or the release of a toxic hazard in the building. Sometimes it is, but more often it is not. A differential diagnosis of building-associated outbreaks is outlined in Table 11.1.

**Table 11.1.** Differential Diagnosis of Building-Associated Outbreaks

- 
1. Building-associated disease
    - 1.1. Hypersensitivity pneumonitis
    - 1.2. Legionnaires' disease
    - 1.3. Infection spread by droplet nuclei (tuberculosis, influenza)
  2. Allergy, associated with presence of a specific allergen
    - 2.1. Mold
    - 2.2. Indoor plants
    - 2.3. Other
  3. Low humidity
  4. Indoor air quality problem (tight building syndrome), unrecognized
  5. Indoor air quality problem (tight building syndrome), with psychogenic overlay
  6. Exaggerated response to odor, with vasovagal reaction
  7. Psychogenic outbreak
    - 7.1. Anxiety
    - 7.2. Sublimated dissatisfaction over work conditions
    - 7.3. Response to psychosocial stressors at work (see Chapter 13)
    - 7.4. Deliberate provocation
  8. Community illness, manifested in a small group in building
  9. Community illness, manifested by a large fraction of the workers in the building
    - 9.1. Diarrheal disease (consider waterborne or foodborne disease)
    - 9.2. Airborne disease (usually respiratory tract infections)
- 

Collectively, incidents of illness among occupants of a particular building are called “building-associated outbreaks.” This is a generic term. When only individuals are affected, this chapter will use the term “building-associated illness.” Both outbreaks and individual cases are very common.

The investigation of building-associated outbreaks is usually undertaken by occupational (industrial) hygienists. The physician responsible for a given workplace may be called in first or as a team member in collaboration with the hygienist. Physicians in community-based practice usually see individual workers in the clinic and are often asked to follow up with more extensive investigations.

Unfortunately, the physician is usually called in too late to do the single most important medical investigation that can be done, which is a symptom inventory and onset survey. It is therefore wise for the physician who is likely to be called in such a situation to put together a kit in advance, and to notify his or her clients or employer that he or she is prepared to investigate these cases. When the call comes, the essential investigation can then be done with minimal delay.

The “sick building syndrome” (SBS) describes a stereotyped but non-specific outbreak in a building, the cause of which is not clearly identified. SBS involves complaints of mucosal irritation, headache, and difficulty concentrating, among other symptoms. It almost always affects a minority of workers in a building, rather than the majority. Ruling out other patterns of illness, and confirming that the pattern under evaluation is a “typical” SBS, is an important branch point in the investigation of a “sick building” because it narrows the diagnostic possibilities and points in a certain direction for evaluation.

The implication of the term “sick building” is that a characteristic of the building is making workers sick, although that is sometimes not the case. The building may have unfavorable characteristics, such as an inadequate, poorly designed, or badly managed ventilation system. Other unfavorable characteristics may include an unrecognized point-source of chemical emissions in the building, distributed sources in the building (such as smoking), dust, dampness, very low humidity, crowding, odor, or any of a number of ill-defined factors that make people uncomfortable. A “sick” building, however, may get its reputation as much from rumor as from validated incidents of illness. Once a building is labeled as a bad place to work, reports of illness multiply and may continue long after the problem, if any, is corrected.

The “tight building syndrome” (TBS), sometimes called the “closed” building syndrome, refers to the accumulation of indoor air pollutants due to decreased ventilation and air turnover. This may have resulted from efforts to seal and insulate buildings in an energy-efficient manner. Energy conservation measures, construction materials, and ventilation practices in vogue from approximately the 1960s through the 1990s, have sometimes resulted in the accumulation of

airborne contaminants to levels that may cause symptoms in susceptible persons. This has been true particularly of the first wave of "green" buildings in the 1980s, which were sealed too tightly in order to maximize energy efficiency. Architects and builders are now much more aware of the problem.

A significant proportion of these outbreaks is psychogenic and reflects deeper psychosocial issues in the workplace (see Chapter 13). Psychogenic outbreaks can be positively identified by certain characteristics or by exclusion, by ruling out air quality problems. The conscientious OEM physician will be aware of the possibility of psychogenic causes in a building-associated outbreak with a compatible profile, but will always consider physical and chemical issues, as well as air quality, as possible causes.

Distinguishing between outbreaks of non-specific illness associated with inadequate indoor air quality, the "tight building syndrome" (TBS), outbreaks arising sporadically due to community illness, and those arising through psychogenic suggestion is the major challenge facing occupational health professionals assessing these problems. The challenge will probably continue until the current inventory of office buildings from that era is replaced or refurbished.

One approach to evaluating building-related illness is outlined in the next section. A more general discussion of building-associated outbreaks then follows in later sections.

## **TRIAGE OF BUILDING-ASSOCIATED OUTBREAKS AND ILLNESS**

There are many approaches to the evaluation of building-related illness. In the past, the mainstay has been exhaustive occupational hygiene evaluations. However, such studies are very expensive and often inconclusive. Many building-associated outbreaks are resolved by non-specific measures (such as manipulating the ventilation in the building) or disappear over time, although not without some residual resentment on the part of the affected workers. Clearly, more productive and robust approaches are needed. There are many studies of sick

buildings, mostly as case studies, but little firm guidance on the best practices to implement. This is in part because of the inherent difficulty in conducting research with consistent protocols on the scale of multiple buildings, which may have different owners.

This section outlines a general approach to triage and initial assessment of buildings for the OEM physician, who normally works as part of a team including an occupational hygienist. The sequence is:

1. Rule out the possibility of the building's posing a serious, life-threatening health hazard.
2. Assess the carbon monoxide level.
3. Check the air exchange, temperature, and humidity, and determine when recent changes were made, if any, in the building (including renovations).
4. Conduct a symptom inventory and onset survey. Analyze for patterns of symptoms and for evidence of either common source transmission (characteristic of air quality issues) or “person-to-person transmission” (characteristic of psychogenic outbreaks)
5. Initiate a limited occupational (industrial) hygiene investigation emphasizing ventilation, carbon monoxide (revisited because it may be intermittent), volatile organic compounds, dust, and odors. Normally, this would be done simultaneously with Step 4.
6. If the above measures do not identify a root cause, proceed to a comprehensive hygiene investigation.

Steps 1 and 2 are essential to ensure the safety of the workers occupying the building. Item 3 is a simple check for problems that are easily corrected. Step 4 is critical to confirm whether the outbreak is consistent with the generalized SBS, or has other features that should take the investigation in a different direction. If the outbreak is compatible with SBS, the critical step is then to determine

if the outbreak is associated with ventilation and poor indoor air quality, or if it may be psychogenic in origin. Step 5, the hygiene survey, should be undertaken by a hygienist with a general, overall understanding of indoor air quality issues, rather than by a technician with specialized expertise in one area (such as mold or asbestos). If there is reason, on the basis of these steps, to suspect a chemical or biological hazard, then the hygiene evaluation can be scaled up to more comprehensive action, in Step 6.

Health assessment methods available to the OEM physician can allow a fairly rapid triage of building-associated outbreaks and a separation of those related to the indoor atmosphere from those likely to have other causes. This depends on a rapid response in gathering data, however. It is of utmost importance to conduct an inventory of the symptoms as quickly as possible, and to capture information at the onset of symptoms. An experienced occupational hygienist should be called in to evaluate the building, following a realistic protocol of investigation. However, the tools available to assess this problem are not limited to occupational hygiene and exposure assessment, as valuable as those may be.

The classic symptoms of SBS, enumerated below, reflect four underlying processes: mucosal irritation, distraction, fatigue, and anxiety over what the symptoms mean. Symptoms that are not consistent with these four processes (and that fall outside the typical, stereotyped syndrome described below) suggest that the outbreak is not a typical SBS and needs to be investigated on its own terms.

### **Building-Related Disease**

An acute respiratory illness that is life threatening or associated with high-grade symptoms requires rapid investigation, of course. If the affected worker agrees, the investigation should be carried out in consultation with the treating physician and with public health authorities.

If one or more cases of pneumonia are diagnosed as Legionnaires' disease or Pontiac disease, there should be a thorough investigation of

the water and cooling systems for the *Legionella* species, which can be difficult to find. An outbreak of tuberculosis may require a screening of co-workers who were in close proximity in the workplace.

Hypersensitivity pneumonitis is a chronic disease (discussed in Chapter 17) arising from an immune response to persistent inhaled antigens. It is often associated with exposure to mold, bird droppings, wood dust, or other dusts that are disturbed and rendered airborne when old buildings are renovated. Initially, hypersensitivity pneumonitis resembles a non-specific, lower respiratory tract illness, but it does not improve with antibiotics and progresses to fibrosis unless exposure stops.

Contact with bird droppings during renovation, which goes beyond the usual scope of building-related outbreaks, may place the worker at risk of cryptococcosis, a very serious, deep mycosis that may present as a pneumonia or meningoencephalitis. Individuals who are immunosuppressed or HIV-positive are at increased risk for this fungal disease. There are many other disorders, including occupational asthma and silicosis, that are associated with construction but would not normally be a problem among office workers.

Chronic, low-grade symptoms other than the SBS syndrome should be evaluated on their own terms. An important sign that there may be a specific illness involved, rather than the generalized SBS syndrome, is the presence of specific, recurrent symptoms that are not compatible with four underlying processes: mucosal irritation, distraction, fatigue, and anxiety. The presence of symptoms such as diarrhea, severe cough, productive cough, chest pain (not burning), and vomiting (not nausea alone) suggest that there may be a specific disorder involved, either associated with the building or arising from the community from which the workers come.

If these symptoms are experienced by a majority of workers in the building or a large fraction distributed in different spatial areas, the root cause is much more likely to be a common source or to be a community-acquired illness. The latter is usually obvious during influenza season, and simply reflects epidemiological trends in the community.

If there is no disorder with compatible symptoms making the rounds in the community, the OEM physician should consider common-source exposure in and around the building, such as surface contamination (for example, norovirus if explosive diarrhea is the problem), the food service serving the workers, or an air quality problem involving the entire building.

Most physicians and other health professionals are poorly prepared in dealing with cases suggestive of toxic inhalation (discussed in detail in Chapter 17) and other responses to major chemical exposures, such as chlorine leaks (which are the most common chemical releases). Such cases are usually obvious, however, and it is unlikely that a true toxic exposure would show great variability in reported symptoms.

### **Temperature and Humidity**

Temperature or humidity that is out of range may result in discomfort for many workers in the building. Temperature or humidity that is within range may still result in individual complaints when it is outside the worker's individual preference range and there is no control over the local environment (for example, if windows cannot be opened). Low humidity is most often associated with building-associated health complaints in office buildings. High humidity and dampness is more often associated with problems in houses.

Temperature and humidity inside office buildings is almost always controlled within bands prescribed by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), which sets voluntary standards, largely on the basis of comfort. These standards are widely followed in North America. ASHRAE standards are based on studies that determine what is acceptable to 80 percent of workers who are sedentary or slightly active, and who are clothed conventionally and appropriately for North America. Individual workers may have other preferences. Workers who are working strenuously in an office environment may also find the range unsuitable. As a result, there will always be some workers who do not feel comfortable in buildings that meet ASHRAE standards but do not

allow individual control of the worker's environment. ASHRAE recognizes two "zones" of environmental control, for summer and winter. ASHRAE standards have been extensively revised since 2000. Standards for relative humidity and temperature are linked by a series of tables. In general, higher temperatures should be accompanied by lower humidity for comfort.

The ASHRAE standard for relative humidity does not specify a minimum humidity; the maximum humidity is 65 percent for occupied spaces, in buildings with dehumidification systems. This standard, adopted in 2007, is significantly below the previous standard, which ranged as high as 80 percent.

Because office buildings are managed to keep the humidity low, the humidity in office workplaces is rarely allowed to become so damp that they become uncomfortable. Excessive moisture is usually not a problem in occupied areas of office buildings. Moisture may accumulate in service areas, behind drywall, in air vents, in drain pans, around windows, above ceiling tiles, and in building cavities. This moisture comes from water leaks and condensation into poorly sealed foundations or basement areas, and by intrusion from the outside. Unobserved spills, leaks in water lines (for example, to drinking water fountains), and precipitation brought in from outside can cause persistent dampness in carpets.

Homes, especially, may be very damp, and have many sources of moisture in bathrooms and kitchens that promote condensation. When persistent moisture is present, the usual problem is mold growth.

Excessive humidity is known to aggravate respiratory symptoms, but the mechanism is not understood. Dampness does not require mold growth or an increase in airborne bio-aerosols for this to occur.

Low humidity is associated with nosebleeds and dry skin, resulting in a prickly sensation on the skin. For this reason, Canadian standards have set a minimum relative humidity of 30 percent. Even so, low humidity is a perennial problem in buildings in cold climates and is often associated with complaints.

Temperature is more obvious to workers than relative humidity, and is also a highly individual preference. For many years, ASHRAE

declined to set a fixed standard, preferring complicated tables for the interaction of temperature and humidity with standards based on dew point, but building operators generally sought to keep offices close to 22.2°C (72°F), which was thought to be optimal. The temperature range now recommended by ASHRAE is 20.0–23.3°C (68–74°F) in the winter and 22.8–26.1°C (73–79°F) in the summer.

Excessive warmth may make workers feel sluggish and distracted. However, with increasing concern over energy efficiency, temperatures in office buildings are allowed to vary more than in the past.

### **Carbon Monoxide**

The cardinal symptom of carbon monoxide toxicity at low levels is headache, followed by visual and neurocognitive changes (see Chapter 10). These symptoms would not be due to carbon monoxide in a well-maintained building with no combustion source. Respiratory symptoms are not seen with carbon monoxide toxicity.

Carbon monoxide may accumulate in buildings to levels that could cause a headache because of intrusion from a parking structure, a covered loading dock, unauthorized use of an air compressor without adequate ventilation, poor ventilation from a kitchen with a gas stove, or open combustion sources such as fireplaces with an inadequate air supply. Because carbon monoxide is heavier than air, it is more likely that a source will be on the same level or above the location where complaints are reported than below.

Carbon monoxide is the only highly toxic agent with no warning signs that would be likely to occur inside an office building with no exceptional features. It is also so hazardous that ruling out exposure is an important safety measure.

### **Radon**

Radon is primarily a problem in homes, not modern office buildings, but may occur in older structures used for any purpose, as well as in poorly ventilated basements of buildings. Radon is a highly variable

hazard. Regions with particularly high levels on average include the Canadian Shield, the Rocky Mountains, and a geologic formation known as the “Reading Prong,” where there is discontinuous surface exposure of pre-Cambrian rock in New York, New Jersey, and eastern Pennsylvania. However, even in regions where there are large expanses of exposed bedrock that can serve as a source, radon levels in individual houses may vary considerably over a small area or even between neighbors. As a gas that is heavier than air, radon tends to accumulate in basements if the air is undisturbed.

Radon ( $^{222}\text{Rn}$ ) is an  $\alpha$ -emitting radionuclide (see Chapter 9), uniquely an inert gas, that forms from the disintegration of radium (Ra) on the way to lead, producing two radioactive polonium isotopes along the way:  $^{222}\text{Rn}$  disintegrates to  $^{218}\text{Po}$ , and then with various intermediates to  $^{214}\text{Po}$ . These so-called radon daughters are also  $\alpha$ -emitters and have a long enough half-life to contribute additional exposure. Because it is a gas, it diffuses readily out of material in which radium is present, such as granite and tracks along cracks and fissures. It then dissolves into and is carried by water, where 0.1 percent may escape. Radon enters basements through cracks and poorly sealed joints in the foundation, through pumped groundwater (especially showers, since the spray enhances the volatilization of the gas), along water and sewer lines, and from radium-containing building materials.

The risk of radon and its daughters is known primarily from studies in the uranium mining industry and in potash mining, where high levels occur. Because  $\alpha$  radiation cannot penetrate skin or other tissue barriers,  $\alpha$ -radiation is primarily a hazard when inhaled. It is interactive (synergistic) with cigarette smoking as a cause of lung cancer, in part because it adsorbs onto particles that can be carried into the deep lung (as described in Chapter 2).

Measurement of radon in structures is easily done by passive dosimetry, using either an  $\alpha$ -track detector or a charcoal canister. This adsorbs the radon and measures it for radioactivity after a few days to a week. Commercial testing kits are available for homeowners. Radon carries an electrostatic charge, which also makes it adhere

readily to glass surfaces. Therefore, one way to assess the problem in a home is to determine the radioactivity deposited on glass windowpanes in the basement.

The problem is easily remediated by ventilation with forced air into the basement and fan exhaust. Sealing the foundation prevents intrusion of the gas. Sub-slab depressurization is a method of capturing and exhausting gas under the foundation before it enters the house.

### **Building-Associated Triggers**

A broad class of building-related disorders involves the presence within buildings of triggers for reactive airways disease and asthma. This is a very common problem, much more common than either building-related disease or SBS. Unfortunately, not much is available in the medical literature with regard to this problem, and it is often overlooked by allergists. However, clinical experience suggests that it is the single most important issue in indoor air quality.

The most obvious association between respiratory symptoms and triggers in the workplace occurs when workers with allergies encounter the allergens to which they are sensitized. Of the most common allergens associated with asthma, mold is the class most likely to be encountered in the workplace. (The others are dust mite, cockroach, and plant allergens.) Occasionally, an individual who has been skin tested for allergies will be found to have tested positive for a mold found in his or her workplace (usually *Aspergillus* spp. or *Penicillium* spp.). However, because conventional allergy testing only covers a small number of possible mold allergens and does not always correlate with respiratory allergy, the absence of a demonstrable skin reaction does not rule out sensitivity to a mold in the workplace. Other allergens that may occur in the office workplace include many of the causes of allergen-induced occupational asthma and plant pollens from decorative plants. Because most workers like to personalize their workplace with familiar objects, it is also possible that allergens (such as dog or cat allergens) can be introduced into the office from the home. In general, the amount of allergen present does not have to be high to trigger a response in a sensitized individual.

Confusingly, workers with allergies may also have their symptoms triggered by exposure to non-specific irritants. These substances do not act as allergens, by triggering an immune response, and may not even be chemically reactive, although some are. They include such common exposures as cigarette smoke, solvents, and dusts. The symptoms are similar to low-grade respiratory infections but are due to irritation of the airways. Occasionally, individuals may have an acquired susceptibility to irritant effects on mucous membranes as well. This is common for a time (sometimes weeks) after an acute respiratory tract infection. However, the effects of non-specific respiratory triggers are much more commonly seen among workers with atopy.

Atopy is very common, seen in about 12 percent of the adult population in North America. It is an inherited condition predisposing the individual to allergies, but it is also a marker for reactive airways and susceptibility to mucosal irritation. Individuals with the manifestations of atopy (multiple allergies, allergic rhinitis, asthma, sinusitis, and eczema in childhood) are much more susceptible to the effects of non-specific airborne irritants. They often have both lower airway reactivity, more often associated with cough and chest discomfort than with discrete episodes of bronchospasm, and upper airway reactivity, associated with sneeze, throat and eye irritation, and voice changes. Occasionally, individuals with low-grade atopy will experience chest tightness and cough in cold weather, but will not be aware of bronchospasm. Some patients have a cough consistent with a low-grade bronchitis, which is sometimes scantily productive and often seasonal.

Atopic individuals have airway reactivity, but only those with obvious reversible airflow obstruction are diagnosed with asthma. As a result, most people with the condition do not consider themselves to be asthmatic; or, if they wheezed as a child, they may consider themselves to have “grown out of it.” Lower grades of atopy, such as in workers with only seasonal rhinitis, may not experience wheezing at all, or may have it only during respiratory tract infections. Low-grade asthma symptoms, allergic bronchitis, or “borderline” findings on

pulmonary function tests are also on the continuum of this condition. Such people will have a positive methacholine challenge, but this test is unnecessary for diagnosis unless there is some medicolegal reason to confirm the condition.

Low-grade neurocognitive symptoms are common in people with atopy, and often accompany aggravation of atopic symptoms. Most often they express themselves through fatigue, malaise, poor short-term memory, inattention, easy distractibility, impaired problem solving, and sleepiness on the job. Many patients use the colorful expression “brain fog” to describe a state of poor concentration and receptivity to information. There are many explanations why a patient with atopy would have this condition, which depend on the patients’ individual circumstances. Medications for allergy, especially antihistamines, may be associated with drowsiness. The inflammatory component of allergy, particularly in cases of sinusitis, may be associated with cytokine release and malaise by the same mechanism as in a case of influenza or a bad cold. Symptoms of allergy are very distracting, particularly if accompanied by vasomotor rhinitis and nasal congestion. However, the most common root cause for neurocognitive symptoms in these people may be sleep disorder. Inability to sleep through the night and poor-quality sleep is associated with markedly reduced short-term memory, distractibility, and difficulty with problem solving. This may be the major reason why workers with allergy, alone, have dramatic decrements in self-reported productivity when they are experiencing symptoms at work. On the other hand, the “brain fog” phenomenon is hardly rare among people who are distracted, anxious, stressed, or fatigued, or have trouble sleeping for other reasons.

Coughing and sneezing are normal defense mechanisms of the respiratory tract. The threshold for response is lower than in individuals without a history of atopy. Atopic people usually cannot tolerate sidestream cigarette smoke, and they cough or sneeze around strong fragrances, experience eye or nose irritation when they use marking pens, often have symptoms when they pump their own gasoline, and experience discomfort in dusty environments. Teachers with the

condition often report cough and sneezing after inhaling chalk dust from blackboards. Paper dust and other otherwise benign “nuisance dusts” are sufficiently irritating to induce sneezing and cough when inhaled.

The modern office is full of low-grade irritants that may serve as non-specific triggers for respiratory symptoms in such individuals. If housekeeping is not meticulous (which it rarely is in the modern office building), dust may accumulate near the worker’s breathing zone, to be inhaled when disturbed. Ozone is generated by some copiers, although manufacturers have now controlled emissions for the most part. Electrostatic attraction may draw dust to screens and video display terminals (apparently less often to LED screens). Volatile organic compounds are less often a problem in the modern office, now that “white-out” is less often used and no longer contains toluene. Nail polish and polish remover (acetone) and spray cleaners are common sources.

Symptoms are often reported by susceptible workers during building renovation, maintenance, or repair work. This is not a surprise because even limited workspace renovations tend to be dusty. Removal of ceiling tiles and drywall opens previously closed-off spaces and allows dust and allergens to reach the breathing zone of workers. Volatile organic compounds may be released by adhesives, new carpet, textiles, and upholstered furniture. In the past, urea-formaldehyde insulation released airborne formaldehyde, and this remains a problem in some low-cost housing situations (most recently in mobile homes) but is no longer an issue in offices.

Cases involving building-related triggers usually present as individuals, or occasionally in pairs when one person hears about or seeks out another person with a similar affliction. In general, they do not appear in large numbers in a building. Because of this, supervisors and human resources management often question the legitimacy of the complaints and, usually after an initial period of cooperation, try to deal with the problem by getting rid of the worker.

## The Sick Building Syndrome

If there are no features of the disorder that suggest a higher level of risk, the next step is to confirm that the outbreak represents the sick building syndrome. SBS is not in itself a diagnosis, and confirming that the symptom pattern is compatible with SBS is not the end of the evaluation. It is only the beginning of another branch in the evaluation, in which a further assessment needs to be made as to cause, choosing between indoor air quality and psychogenic factors.

The symptoms reported by occupants of a “sick” building are stereotypical but non-specific. They include headache, eye and throat irritation, fatigue, mild (very occasional) cough, difficulty focusing attention, chest tightness or burning (often subjectively causing alarm but objectively unlikely to be profound shortness of breath), and a general feeling of being “unwell” while in the building. These symptoms are consistent with four main underlying conditions: mucosal irritation, distraction, fatigue, and anxiety provoked by worry over the significance of the disorder.

The cardinal symptoms of SBS are not life threatening but may be very uncomfortable and may interfere with productivity and satisfaction on the job. They do provoke anxiety and may be associated with distraction, and with discomfort and sleep disorders, as noted above.

These symptoms are also common symptoms of upper respiratory tract infections, sleep disturbance, and very common chronic conditions associated with atopy, such as mild asthma, allergic rhinitis, and sinusitis. Atopy confers susceptibility to the irritant effects of dust, volatile organics, and bio-aerosols, whether or not their action is sufficient to trigger a response in airways or sneezing. Some of these symptoms (fatigue and headache) are also compatible with carbon monoxide toxicity. The mere reporting of these symptoms, therefore, does not rule an air quality problem out or in. Analysis of data at the time of the onset of these symptoms, however, may rule in psychogenic outbreaks, as discussed below.

Timing is a useful clue only if there is an anomaly and the pattern does not fit the expected pattern of aggravation at work. Most

workers feel better at home for most health problems, and improvement with time away from work is not specific as to cause. Improvement away from work, over weekends, and on vacation, while confirming that there may be a relationship to work, does not really help to separate indoor air problems from other causes. Likewise, the return of symptoms on Mondays or post-shift does not add much useful information. However, if the worker reports symptoms that are not consistent with these patterns, this finding may be evidence that the problem lies outside the workplace or is not SBS. If symptoms are occurring in the middle of the night, for example, the worker may be developing occupational asthma of the delayed variety, which is distinct from SBS.

SBS is a syndrome, a cluster of symptoms and findings, that leads the physician down a certain path of investigation and narrows the possibilities to a more restricted differential diagnosis. The two most important elements of the differential diagnosis are inadequate ventilation with resulting poor indoor air quality, and psychogenic outbreaks. They can be separated by a timely symptom survey or by an occupational hygiene survey. The components of a hygiene survey are discussed in greater detail later in this chapter.

In general, only a small fraction of workers complain of an indoor air quality-related problem at first, unless the exposure or the psychogenic factors are overwhelming. This phenomenon is often mistaken for collusion among those who complain first. As noted, people vary in their susceptibility to irritating exposures and to odors. Usually, only the most susceptible will complain first, and there will usually be a small number of workers who never complain, even under conditions that are intolerable to the majority. Most of those who complain early are unusually susceptible and many will have atopy, which is expressed in the form of reactive airways, chronic upper and lower airway inflammation, sinusitis, and other effects that render the person much more susceptible to non-specific irritants (such as bio-aerosols or volatile organic compounds) than the majority. Those who first notice the problem are often accused of being psychologically suggestible or self-serving. They often adopt

a defensive attitude in response, furthering separation from their co-workers and conveying the impression that they are complainers.

### *Assessment by Health Survey*

Although SBS symptoms are not specific and together only constitute a syndrome of feeling unwell, it is very desirable that a survey of occupants be done as quickly as possible. Assessment of a SBS is achieved by a symptom inventory and onset survey, but this must be done quickly to be reliable. Delay may result in the introduction of recall bias and contamination from subjects hearing the stories of other workers. For prominent “sick buildings,” media coverage may also influence the history given by individuals.

In the greatest detail possible, this survey should record the salient symptoms and—critically—their date, time, and location of onset, for every individual in the building who reports symptoms, not just those in the area most affected. The location where the worker usually works and where he or she first noticed symptoms should be plotted on a floor plan or map. The information will be put to good use. The symptom inventory will either confirm that the outbreak is an SBS, or identify anomalous symptoms that require investigation. (Some background “noise” in reported symptoms is to be expected, but these will usually be consistent with the inventory of preexisting conditions.) The symptom onset data can be used to map an epidemic curve, to determine whether the pattern of transmission is “common source” (suggesting an air quality problem or a common source in the building, such as a food service) or propagation “person to person” (suggesting psychogenic factors). The spatial relationships will be used to confirm the pattern of transmission for groups, and to indicate where in the building occupational hygiene investigations should be most intensive.

The symptom inventory and onset survey can be conducted by questionnaire or interview. Each worker should complete the survey instrument or be interviewed alone, apart from the other workers in the building. The inventory should always capture detailed information about when an individual’s symptoms began, because

this information will be important in the next phase of the evaluation. Having the worker mark on a calendar the day that he or she first developed symptoms is useful. The inventory should extend beyond the usual respiratory and neurocognitive symptoms. In a neutral manner, as in a review of systems, the questionnaire or interview should give the worker the opportunity to report gastrointestinal, dermatological, and many other types of symptoms that are not part of the SBS. It should also capture information on preexisting conditions, particularly allergy and asthma.

If typical SBS symptoms are present among a substantial fraction of workers housed in a particular area (roughly 10 percent or above, but no threshold has been empirically confirmed), the problem is compatible with SBS. If the reports of symptoms or feeling unwell are limited to only a few individuals, the event is probably not a true building-related outbreak. It more likely reflects individual health issues or responses to local indoor environment, such as allergens.

There should be at least a few open-ended questions about occupants' concerns and whether they wish to receive information on the investigation. In addition to capturing essential information for evaluation, the survey communicates clearly that management is concerned about the health of workers and allows workers to express their concerns. It is also an invaluable aid to risk communication as the evaluation progresses.

### *Assessment by Hygiene Survey*

A walkthrough of the building by the entire team is, of course, essential. The evaluation should be quick but thorough. Common sources of indoor air quality problems should be emphasized. But all parts of the building that are either occupied or that communicate with occupied space should be inspected. If a problem is found, it should be noted and assessed, but the assumption should not be made that the first or the most obvious defect encountered is necessarily the cause of the problem. Table 11.2 lists common sources of indoor air problems in office buildings.

**Table 11.2.** Common Sources of Pollution Inside Buildings, with Examples

- 
- Ambient air pollution: via ventilation intake or windows
  - Animals: allergens
  - Building materials: urea formaldehyde insulation, asbestos insulation or ceiling tiles
  - Fires or fireplaces: carbon monoxide and hydrocarbons
  - Furnishings and textiles: out-gassing from fabric sizing, new carpets
  - Industrial processes: occurring inside building
  - Kitchen exhaust or smoke: from cafeterias, restaurants
  - Loading area: diesel exhaust
  - Maintenance activities: cleaning solutions, rug shampoo, dust
  - Parking structures: carbon monoxide
  - Plants and gardens: allergens
  - Tobacco smoke: smoking by workers, visitors
  - Ventilation system: contaminated humidifier systems, air conditioning
- 

Occupational hygienists should perform a screening evaluation of the building, emphasizing ventilation, carbon monoxide, volatile organic compounds, the presence of mold, and dust. If there is no specific problem with indoor air quality, and no indication on inspection of an unusual source, this is usually sufficient. If there is a realistic possibility of an uncommon chemical exposure, methods that are more elaborate may be required, as described in the next section.

A comprehensive investigation of a building problem would proceed as outlined in the next section, but at the outset, on a preliminary basis, it should address four critical concerns:

- Intake of air into the building. For example, is the air intake close to an emissions source or drawing exhaust air from the building? Is there a loading dock that acts as a portal for carbon monoxide?
- Unusual activity outside the building (for example, repaving of asphalt driveways or parking lots)

- Problems with the cooling, humidification, and dehumidification systems
- Sources within the building. For example, have there been any changes in equipment or activity? Is a kitchen being used? Are repairs or maintenance being done in the building? Have new furnishings been introduced?
- Ventilation (air exchanges, carbon dioxide level)

Most SBS outbreaks involving air quality are relatively localized. Their location may represent points of imbalance in the ventilation system.

Initial evaluations of building-associated outbreaks should emphasize a qualitative and semi-quantitative hazard evaluation. An inspection should normally reveal sources of chemical release or indicate sources of odor. Sampling should be undertaken, depending on the circumstances of the exposure, in the areas most affected, in ventilation ducts, at plausible sources, in parking structures, and in areas of special or unusual activity.

The ASHRAE recommended air exchange rate, recently revised upward, is  $0.567 \text{ m}^3/\text{min}$  (20 cubic feet per minute) for office buildings with no unusual activity or smoking allowed. (Smoking lounges may require three times as much flow.)

Later, if no plausible cause is discovered and the pattern is not compatible with psychogenic factors, a more comprehensive occupational hygiene survey can be done, as described in the next section.

### **Identifying Incidents of Psychogenic Origin**

Incidents of psychogenic origin may reveal themselves directly if information on the onset of symptoms can be determined accurately from individual workers, and without contamination from other workers and the media. This requires quick action and the collection of data by questionnaire or interview as soon as the outbreak comes to attention. It is important to obtain the data before there is extensive

discussion among workers in various parts of the building, potentially contaminating the self-reported findings. In general, psychogenic outbreaks will show the following characteristics:

- A pattern of transmission from person to person within a spatial unit (such as an office suite)
- A pattern of sequential spread from spatial unit to spatial unit
- Symptom reporting in proportion to the visibility of the problem and, if applicable, news coverage
- An onset of symptoms that does not reflect ventilation and airflow patterns in the building
- Symptoms that are subjective and compatible with the non-specific SBS syndrome (feeling unwell, malaise, throat and eye irritation, difficulty concentrating)
- Presence in the work environment of a basis for heightened worker awareness or anxiety over their health (illness or death of a co-worker, identification of asbestos in the building, intermittent presence of an odor)
- A triggering event (this may be an unfamiliar odor, collapse of a ceiling tile, the acute illness of a co-worker, the discovery of mold, or some other precipitating event)
- Complaints of symptoms that begin before the worker actually enters the building

The conclusion that a building-associated outbreak is psychogenic is certain to be disputed. The problem is complicated by the clear demonstration of psychogenic factors acting in some outbreaks clearly related to poor indoor air quality, and the frequent role of odors in triggering psychogenic outbreaks. With increasing attention being given to the “tight building” phenomenon by the media, incidents in which the cause is psychological rather than chemical may increase in frequency due to suggestion.

## Chemical Hazards

Lay persons often assume that a building-associated outbreak is caused by an unrecognized TBS with an overlooked or unusual chemical exposure not detectable by conventional means. Experience suggests that this is exceedingly rare. There is rarely a substance "X," undiscovered and unrecognized, that significantly affects health. Nevertheless, the occasional recognition of a new hazard suggests that this possibility can never be completely ignored.

Truly dangerous unrecognized toxic hazards encountered in the context of a TBS must be exceedingly rare, outside of obvious high-risk settings. When toxic agents released within buildings accumulate, or there is contamination of the system by infectious agents or antigens, buildings with inadequate ventilation systems may be implicated in unusual outbreaks of serious acute or chronic illness. Such toxic agents are obviously very uncommon in office buildings but with contemporary concerns over terrorism and malicious intent, and the rare instance of unauthorized activity within the building or migration from elsewhere, the possibility cannot be dismissed. The most obvious candidate for a toxic exposure in a building would be hydrogen sulfide from sewer gas, but this would normally be obvious from the odor.

A common irritant exposure in building settings is formaldehyde. Exposure to formaldehyde is discussed in the next section.

## Odor

An odor that is nauseating, unexpected, overwhelmingly strong, or linked psychologically to unpleasant associations (such as body odor or feces) may precipitate nausea, lightheadedness, and a vasovagal reaction in susceptible individuals. Even barely detectable odors may cause anxiety or concern among workers with well-developed olfactory senses who are fearful of their health or who have past association with the odor. Certain odors, such as that of acrid perspiration, burning plastic, or feces, carry such strongly negative

psychological associations that they may make people ill even when barely detectable to others.

The appearance of an unexplained odor, or the recurrent appearance of a foul odor, often leads to the conclusion that something is fundamentally wrong with the building. This is particularly true for drainpipe odors, which may be interpreted as dangerous sewer gas, and dank, earthy smells, which raise fears of mold contamination and mycotoxin toxicity. Drainpipe odors commonly result when the standing water in the U-trap of the drain evaporates and allows the passage of gases back into the room. It can be solved by pouring water into the drain. Earthy smells are called “geosmia” and are characteristic of mold growth, being caused by the same volatile mold products that cause the smell in soil.

Individuals vary greatly in the acuity of their sense of smell, which tends to be more acute in women and non-smokers, and also becomes more acute during pregnancy. This phenomenon is thought to be an evolutionary adaptation to protect a person against the ingestion of poisonous plants or spoiled food, similar to the aversion to bitter taste.

## **INDOOR AIR QUALITY**

Buildings are containers for people and goods built to keep the outdoors out. This means that they also keep the indoors in. Unless the communication with the air outside is managed appropriately, airborne dust, volatile chemicals, emissions, and bio-aerosols can accumulate inside the building. Building owners and operators manage the indoor atmosphere through the “heating, ventilation, and air conditioning” (HVAC) system, with the goal of achieving adequate ventilation and a sufficient number of air exchanges to keep air fresh in the building,

The quality of air inside buildings has emerged as a major environmental problem in recent years. The emphasis on energy conservation since the 1973 oil embargo has led to increasingly tightly sealed structures and reliance on ventilation systems to maintain acceptable air

quality. Often, the result has been the accumulation of airborne contaminants and deterioration of the quality of air inside the building.

The tight building syndrome is a real phenomenon that can be measured and evaluated. Indeed, some investigators have suggested that indoor air hydrocarbon concentrations correlate most closely with symptoms usually associated with the classic sick building syndrome, but the association is not necessarily causal; it may simply reflect ventilation. Conversely, generalized, non-specific symptoms among workers are more often reported in buildings with air conditioning than in those with simple mechanical ventilation.

Studies in both residential and office settings have confirmed that the quality of the air inside buildings is often worse than the air quality outside. This is because of sources from within the building. A good HVAC system with adequate air exchange should carry in less pollution than air outdoors, although air contaminants in the outdoor air will track in parallel with indoor air quality. The quality of air inside a building does bear some relationship to the quality of air outside, called the ambient air quality. Generally, outside air enters the building through windows, doors, and especially the ventilation system intake; therefore, whatever pollutants may be present outside will eventually appear inside the building. This relationship is not simple, however. There are many sources of air pollution inside buildings that add to the problem, and these create a qualitatively different profile of indoor air quality. Additionally, the concentration of specific pollutants entering the building with the ambient air may be affected by air conditioning and filtering. If the concentration of the pollutant is changing rapidly in ambient air, the concentration inside the building usually lags behind to a degree that depends on the rate of air exchange.

### **Indoor Air Pollutants**

Table 11.3 presents a list, by no means comprehensive, of common indoor air pollutants and their associated sources within the building. Most of these air pollutants are described elsewhere in greater detail (see Chapters 10 and 12).

**Table 11.3.** Indoor Air Pollutants and Their Common Sources

Allergens	Reconstruction (exposing old interior surfaces contaminated by molds), plants, animals, humidifiers.
Asbestos	Insulation, building materials
Carbon monoxide	Auto and truck parking, open fires
Formaldehyde	Insulation, fabrics and furniture, cigarette smoke
Infectious agents	People, humidifiers
Nitrogen dioxide	Gas stoves, fires, cigarette smoke
Organic chemicals	Cleaning agents, solvents, aerosol sprays, fires, cooking activities, cigarette smoke
Ozone	Electrical equipment, photocopiers
Radon daughters	Stone, groundwater leakage, soil

Formaldehyde remains a significant problem in buildings. Urea formaldehyde foam insulation (UFFI) was heavily used in the 1970s and raised major concerns with respect to risk. It was banned in Canada in 1980 and by the U.S. Consumer Product Safety Commission in 1982, but the ban was challenged in court and overturned on appeal, on grounds that the evidence for risk was weak. Notwithstanding this history, UFFI was effectively banned by the U.S. Department of Housing and Urban Development (HUD) twice—first in 1983, when it prohibited use of UFFI in low-cost housing that qualified for HUD-sponsored mortgage insurance, and again in 1985, when a stringent limit (0.1 ppm) was set on indoor exposure levels in projects funded by HUD. Subsequently, the market for UFFI weakened and alternatives became more popular. In 2004 the International Agency for Research on Cancer determined that formaldehyde was a human carcinogen.

Although indoor levels have dropped overall, formaldehyde remains a problem in certain situations (most recently in the United States in temporary mobile homes known as “FEMA trailers” which were supplied to displaced evacuees after Hurricane Katrina in 2006). It is also a problem in the form of emissions from products brought into the building. The major source of formaldehyde in most

office settings today is processed wood products (such as oriented strand board) in furniture, paneling, building materials that may be part of renovations, adhesives (especially adhesive-backed tiles) or cigarette smoke, if smoking is allowed in the building. The ASHRAE standard for formaldehyde is 0.1 ppm, much lower than the OSHA standard of 0.75 ppm 8-hour TWA, which is intended for industrial workplaces in which formaldehyde is present. However, the ASHRAE standard is higher than the World Health Organization guideline of 0.08 ppm.

Emissions from furniture have increased as a problem because of changes in manufacturing techniques and shipping. As solid wood has become more expensive, more furniture is made of particleboard cores with a wood veneer. Formaldehyde out-gassing from the finished product is then trapped within the plastic wrapping used to protect the furniture during shipping. The furniture manufacturing sector, like most manufacturing, now operates on a “just in time” production basis, in order to maintain low inventory costs. As a consequence, the furniture does not have time to outgas, since the accumulated emissions do not disperse during storage. Volatile organic compounds are released from coatings on furniture and sizing of the fabric. When the product arrives at the loading dock of an office building, it may be emitting high levels of formaldehyde, solvent-related volatile organic compounds (such as xylenes), terpenes (volatile organic compounds released from wood, including pinenes), wood-derived derivative products (aldehydes that appear to be produced during the heating of wood), and other gases (including trichlorofluoromethane, source unknown). These are released in high concentrations over a short period of time when the plastic wrapper is removed. Out-gassing is accelerated in furniture that is placed in a warm room, in the sun, or near a radiator, where it is heated. Brominated fire retardants, which bio-accumulate, are also emitted from furniture. In 2007 the Building and Institutional Furniture Manufacturing Association received approval from the American National Standards Institute to make its emissions standards and testing protocol the industry voluntary standard.

Cleaning agents may release volatile organic compounds into confined spaces. Common chemicals used in cleaning agents are limonene,  $\alpha$ -pinene, acetone, xylenes, decane, ethanol, ethylbenzene, and, in small amounts, hexane. These are usually mildly irritating at worst.

Cigarette smoke has been a major contributor to indoor air pollution. The advent of anti-smoking bylaws in most American and Canadian cities, and OSHA regulations that protect workers from passive smoke in the workplace, have substantially improved the situation. Smoking is still tolerated in some jurisdictions, in some non-office workplaces (such as the “smoking” areas of bars), in smoking lounges, and in close proximity to the entrances to buildings that do not have rules against nearby smoking. Table 11.4 lists some of the more toxic constituents of cigarette smoke. Smoke emissions from burning cigarettes are called “sidestream smoke,” and the exhaled smoke inspired by smokers is called “mainstream smoke.” “Sidestream” smoke is of greatest concern. Mainstream smoke, as inhaled by the smoker, is filtered through the cigarette itself and is further modified by the respiratory tract of the smoker: this reduces the risk to passive smokers, although the risk to smokers remains very high. Sidestream smoke, on the other hand, has a much higher concentration of major toxic constituents, including several recognized carcinogens. Sidestream smoke is further modified by dilution with room air, and possibly by chemical and odor changes over time after emission, a process called “aging.” This process, however, is still poorly characterized by chemical analysis. Inhalation of both sidestream and mainstream smoke by a non-smoker who happens to be present in the environment is called “passive smoking.” Exposure to constituents of cigarette smoke that adhere to surfaces, such as table tops and windows, or that deposit on textiles has been called “tertiary smoking” and is a particular concern for exposure of children.

Asbestos remains a concern in many buildings but is not associated with building-associated outbreaks. The health effects are chronic and do not contribute to the SBS. The hazard of asbestos in buildings is by now almost entirely limited to workers who remove asbestos insulation.

**Table 11.4.** Constituents of Sidestream Cigarette Smoke

Vapor Phase	Particulate Phase
Acetic acid	4-Aminobiphenyl
Acetone	Anatabine
Acrolein	Aniline
Ammonia	Benz[a]anthracene
Benzene	Benz[a]pyrene
Carbon dioxide	Benzoic acid
Carbon monoxide	$\gamma$ -Butyrolactone
Carbonyl sulfide	Cadmium
Dimethylamine	Catechol
Formaldehyde	Cholesterol
Formic acid	Glycolic acid
Hydrazine	Hydroquinone
Hydrogen cyanide	2-Naphthylamine
Methylamine	Nickel
Methylchloride	Nicotine
3-Methylpyridine	N-Nitrosodiethanolamine
Nitrogen oxides	N-Nitrosonornicotine
N-nitrosodiethylamine	Particulate matter
N-nitrosodimethylamine	Phenol
N-nitrosopyrrolidine	Polonium-210
Pyridine	Quinoline
Toluene	Succinic acid
3-Vinylpyridine	2-Toluidine
	Zinc

### Investigating Building-Associated Outbreaks

Systematic investigation of a building-associated outbreak is best conducted using a team approach involving an occupational hygienist, a health professional, and the building supervisor, at a minimum. This subsection assumes that the preliminary steps outlined for triage have been performed and have not identified a cause.

The first step in the investigation of a building-associated outbreak is to return to the time and date of the onset of reported symptoms,

and the precise location in the building where the affected workers spend their time, reviewing the information captured in the survey. This information should point to a particular location within the building, either on the basis of a source of exposure or on the flow characteristics of the building's ventilation system.

Two general approaches have been applied to the problem of distinguishing between incidents that reflect significant exposures and those that reflect psychological influences. The first is the exhaustive documentation of health complaints in order to identify patterns and to localize possible sources by inference. Unfortunately, most outbreaks are poorly characterized by these means because the symptoms are non-specific. The second approach is to rule out significant chemical exposure by exhaustive environmental testing, an expensive and often unsatisfying exercise because of lingering doubts and unanswerable suspicions among employees that a significant hazard might have been missed. Although improbable, cases in which a previously unsuspected or unusual toxic exposure is ultimately identified are widely discussed (some of them are urban legends, but there are examples) and lend credence to this skepticism. A more rewarding approach is to trace the distribution of cases in space and time, pointing to either a social or common source transmission. This epidemiologic approach can be used to analyze an incident in which the precipitating events cannot be determined. In all such cases, the absence of evidence for significant chemical exposures should be well documented, which is the job of the hygienist.

### *Ventilation*

The ventilation system should be inspected, with careful attention to the ducting, filters, and humidifiers, as well as inspection of the recent maintenance records. If there is nothing untoward—no suggestion of a chemical exposure, the humidity level is normal, and the pattern does not logically coincide with building ventilation characteristics—then the possibility of psychogenic factors should be considered.

Ventilation is usually evaluated using carbon dioxide measurements. Since occupants of the building produce carbon dioxide, which is removed by the ventilation system, its accumulation is an indication of the inadequacy of the ventilation. Ventilation can also be assessed locally with smoke tubes, which indicate the direction of air flow, and by a velometer, a device that measures air flow out of a duct.

Building ventilation systems are complex mechanically and in their performance characteristics. Essentially, a modern building ventilation system must maintain air at an appropriate temperature and humidity regardless of ambient conditions, and must maintain sufficient turnover to keep the level of pollutants and of carbon dioxide at acceptable levels. The activation of heating systems is usually accompanied by a significant lowering of humidity, particularly if cold, relatively dry air is drawn into the building and there is no humidifier. Start-up of the ventilation system may also cause dispersal of dust accumulated in the heating units and ducts. Low humidity and unwanted dust in the air are well-known causes of eye and upper respiratory tract irritation in homes and workplaces in cold climates.

To achieve the winter standard, ventilation systems should operate efficiently and without excessive intake of cold air. However, “make-up” or replacement air is required to dilute and expel indoor air pollutants. Thus, there is always a trade-off between turnover of air and energy efficiency. In the past, this problem has not been so apparent because most buildings were relatively leaky, allowing free passage of air through doors and windows. Today, however, structures are engineered to be much tighter, sometimes almost hermetically sealed. Unless sufficient make-up air is introduced through the ventilation system, accumulated air pollutants cannot disperse.

### ***Evaluation of Air Quality***

During walkthrough, numerous measurements may be collected using direct-reading devices (Draeger tubes and hydrocarbon analyzers). For more quantitative analysis in areas of concern, investigators

**Table 11.5.** Standard Methods for Evaluating Chemical Exposure in Building-Associated Outbreaks

Asbestos	Sampling pump-filter, microscopy
Organic vapors	Sampling pump: GC/MS portable GC direct reading
Sulfur dioxide	Direct reading meter
Hydrogen sulfide	Direct reading meter
Carbon monoxide	Direct reading meter
Formaldehyde	Passive dosimeter
Chlorine	Draeger system
Dust particulate	Sampling pump, collector
Oil/mist	Sampling pump, collector
Halogenated compounds (including chlorinated compounds, pesticides, fluorocarbons)	Portable GC/GC-MS

collect air samples using charcoal tubes, which should be promptly transported to qualified laboratories for independent analysis using gas chromatography and mass spectrometry (GC-MS). Table 11.5 outlines the usual procedures for common, known exposures.

In complex cases, or those in which an elusive or unknown but presumed hazardous exposure is expected, investigators may add more sophisticated portable instruments, employing direct-reading photo-ionization (H-Na), flame ionization, and electron capture gas chromatography, to the initial analysis. Also, they may repeat the collection of samples for GC-MS, including two-hour samples from air vents and working areas. These studies should be performed during normal ventilation and again after the ventilation system has been shut off for 24 hours.

Occupational hygiene measurements can be taken during normal working hours, to demonstrate the usual exposure levels that workers experience. However, measurements may also be taken after a night or weekend in which the HVAC system is off, in order to obtain a profile of levels under a “worst case” scenario.

## PSYCHOGENIC OUTBREAKS

SBS is a syndrome, a cluster of symptoms and findings, that leads the physician down a certain path of investigation and narrows the possibilities to a more restricted differential diagnosis. The two most important elements of the differential diagnosis are inadequate ventilation with resulting poor indoor air quality, and psychogenic outbreaks.

In many cases, outbreaks of acute illness occur among occupants of buildings that are not tightly sealed or for which no exposure can be implicated. For many, and probably most, of these outbreaks, the root cause is a collective response to psychogenic factors. Such incidents cause great anxiety among all parties concerned and usually end inconclusively or with rejection of the explanation. However, psychogenic outbreaks are well documented and are not uncommon.

Talking about psychogenic outbreaks with workers is made unnecessarily difficult by history and past terminology. Psychogenic outbreaks were often referred to in the past, somewhat condescendingly, as "mass hysteria," which was treated as a diagnosis of exclusion. The term "mass hysteria" is value-laden, crudely connoting herd behavior among human beings and suggesting weak individual character, despite an abundant literature in social psychology documenting that irrational group behavior may occur among both genders and all social classes, intelligence levels, and ages.

Using the tools of occupational hygiene, the role of psychogenic factors is virtually impossible to prove, since the conclusion is usually based on exclusion of known exposures; therefore, there will always be lingering doubt about a mysterious, unidentified "substance X." Periodically, new exposures or physical factors are identified that raise the possibility that non-psychological factors responsible for an outbreak have been overlooked, such as rug shampoo, which was not recognized as a potential problem until relatively recently. Affected workers can never be sure that all alternatives have been ruled out unequivocally, and their anxiety is real and, in context, not unjustified.

Analysis of psychogenic outbreaks as a phenomenon has been exceedingly difficult because of variations in circumstances, populations exposed, and activities within buildings studied. Rarely are two incidents similar enough to permit direct comparison or to be studied by the same investigator using identical techniques. The nature of the incidents results in considerable uncertainty because events in the initial phase of the incident are usually lost to documentation and must be reconstructed from memory. However, advances in the epidemiology of building-associated outbreaks and in the study of psychosocial factors in the workplace (see Chapter 13) have clarified the broad outlines of the phenomenon.

### **Dynamics of Psychogenic Outbreaks**

Discomfort and physical symptoms are a socially acceptable manifestation of dissatisfaction or anxiety in the workplace. Where there is a precipitating event, such as the diagnosis of a co-worker with cancer, or an alarming, unexplained odor, the anxiety and workers' feelings of powerlessness become associated with a seemingly concrete issue and are transferred to the perceived hazard. In circumstances of perceived danger or prolonged stress, a group knit by common experience usually develops irrational behavior in anticipation of danger, not during actual confrontation. Thus, in the case of war, the incidence of psychopathology appears to drop during wartime, but incidents described (properly) as mass hysteria may occur during intervals of tense cessation of fighting, continued civil unrest, or deep personal insecurity. A mass hyper-suggestibility state envelops a credulous population that is concerned about its well-being and may be sent into near panic by the communication of rumors or incomplete information. Psychogenic outbreaks therefore have a tendency to "build up" and then explode, triggered by a seemingly minor event. Unknown is the degree to which medical personnel sometimes contribute to the problem, through overly aggressive treatment or by conveying a sense of alarm that the incident does not warrant.

Psychogenic outbreaks present certain characteristic features that may be instructive. Table 11.6 presents a list of features that suggest psychogenic outbreaks. Of particular importance are the patterns of illness in time and space. The sequences of location and incident cases are incompatible with a point-source exposure, especially one borne by air flow patterns. Instead, the epidemic curve and the reconstructed

**Table 11.6.** Features Suggesting Psychogenic Outbreaks

- 
- A pattern of transmission from person to person within a spatial unit (such as an office suite)
  - A pattern of sequential spread from spatial unit to spatial unit
  - Symptom reporting in proportion to the visibility of the problem and, if applicable, news coverage
  - Onset of symptoms that does not reflect ventilation and air flow patterns in the building
  - Symptoms that are subjective and compatible with the non-specific SBS syndrome (feeling unwell, malaise, throat and eye irritation, difficulty concentrating)
  - Presence in the work environment of a reason workers may have heightened awareness or anxiety over their health (illness or death of a co-worker, identification of asbestos in the building, intermittent presence of an odor)
  - A triggering event (this may be an unfamiliar odor, collapse of a ceiling tile, the acute illness of a co-worker, the discovery of mold, or some other precipitating event)
  - Complaints of symptoms that begin before the worker actually enters the building
  - Shifting location not consistent with ventilation patterns
  - Time sequence not consistent with ventilation flow rate
  - Odor not consistently reported by those affected
  - No detectable chemical exposure
  - Absence of medical findings compatible with exposure
  - Epidemic curve consistent with “person-to-person” rather than “point source” transmission
  - Variability of symptoms reported by physicians, depending on health facility
-

diagram of contacts resemble the pattern of a transmissible agent, a pattern with a remarkably short life cycle and that may be broken by intervention during a subsequent outbreak. What is being transmitted is not an infectious agent, of course, but awareness of the incident and, consequently, anxiety.

Table 11.7 presents features that are probably not reliable indicators of a psychogenic origin of an outbreak. Because anxiety and suggestibility may accompany true exposure incidents, the demonstration of hyperventilation or the presentation of illness among unexposed individuals is unreliable in distinguishing between indoor air quality problems and psychogenic outbreaks. The symptoms presented can include common, non-specific reactions to stress, such as nausea, hyperventilation, or sympathetic symptoms. Reports of odor are strongly prone to suggestion, especially if the odor is described as weak, "chemical" in nature, or transient. Some individuals are more likely to be susceptible to psychogenic factors than others, and these individuals may also be the first to notice and complain of chemical exposures or odors, having a heightened body awareness in general. Serial outbreaks may be due to a persistent low-grade exposure to odor, although this is much less likely if corrective measures are taken, including general housekeeping. The immediate impressions of physicians are probably not very reliable in such situations, in the absence of objective clinical data. Physicians may be biased in their diagnosis by the perceived credibility of the patient, the degree of anxiety experienced by the patient, past experiences of similar complaints, and attitudes toward smoking (when this is

**Table 11.7.** Features That Are Unreliable in Suggesting a Psychogenic Outbreak

- 
- Hyperventilation among affected individuals
  - Immediate reports of odor
  - Psychological profile of the index case
  - Serial outbreaks
  - Immediate response by physicians
  - Illness of obviously unexposed individuals
-

permitted in the building) and toward environmental chemical exposure, especially if they have a limited knowledge of toxicology.

### Typical Scenario

A rather typical scenario for a psychogenic outbreak follows.

The workforce has a history of concern over personal and family health, expressed in confidence to health personnel and through frequent complaints. The concern focuses on cancer and the suspected presence of unidentified carcinogens in the workplace. These concerns have not been fully addressed. Over the period before the outbreak, these concerns intensify. At the time of the outbreak, some tangible but apparently trivial initiating event occurs. An odor may be detected.

The weather may have changed and the heating system may have come on. A foul or unusual odor may be detected by some occupants. Occupants of the building may feel the effects of low humidity, such as itchy skin, eye irritation, a stuffed-up nose, and a scratchy throat. In past years this would have been accepted, but this time their long-smoldering concerns over their health surface. Predictably, this occurs where the largest cohesive group of workers is located and where concern over health among employees is highest.

The first group of workers who declared themselves ill are moved, but the problem soon returns. In the meantime, others with whom these workers have had contact declare themselves ill. The progress of the apparent illness does not conform to ventilation patterns, plant subdivisions, or job responsibilities. It may even spread to adjacent buildings away from the area initially affected. The issue is prolonged, and as it draws out, the workers feel validated in their belief that something is very wrong. Even negative occupational hygiene reports do not alleviate their anxiety. They are perplexed that no cause has been found and begin to suspect a cover-up.

In order to break the cycle of anxiety feeding on uncertainty, and to prepare the workers for interpretation of the occupational hygiene data to come, it sometimes helps to conduct a health education program in the form of open meetings with employees explicitly

addressing their concerns, which often are focused on chemical and environmental causes of cancer. These meetings should be followed by a series of health promotion programs to emphasize continuing concern for the health of employees. This intervention will directly address the most critical contributing factor acting in this incident, and will remove the fuel that could cause this issue to smolder after the occupational hygiene data are announced and no cause is identified.

A health survey to document the problem, conducted by interviews based on questionnaires, is another useful approach. This should be followed by a similar survey in one year to assess progress. The follow-up study is important in order to bring the issue to a definite and recognized conclusion. Because of the therapeutic benefit of reporting one's health condition, it is preferable that all employees and residents take part, not just a sample.

### **Anatomy of a Psychogenic Outbreak**

An unusual opportunity once arose to study a building-associated outbreak in which the cause was clearly psychogenic. The building in question was a modern, one-story, glass, cement, and steel structure occupying most of an urban block in a city in California. On the day in question, it housed 153 employees of a utility, who were engaged primarily in the operation of video display terminals (VDTs), communications consoles, and switching equipment. No products were manufactured on the site, nor were chemicals used to provide the services performed other than in routine maintenance. The building consisted of two sections separated by a central corridor. Ventilation for both sections followed the same pattern: Intake was from the outside where the building fronted a moderately busy arterial street within 0.5 kilometer of a freeway (a source of air pollution) but distant (1.25 km) from the nearest freeway interchange. In both sections the flow pattern was into the work areas and out a central corridor that divided the work areas. Thus the two sections drew from common air sources but their airstreams did not commingle until they were channeled to exhaust. On a midweek spring day in

1984, the index case was at his workstation in the west at 8:00 a.m. when he noticed an odor. He asked to be sent for medical evaluation at 9:40 a.m. due to nausea and was examined by an experienced occupational physician, who found no medical evidence of toxic inhalation. The index case volunteered the information that he was a veteran of the Vietnam War and had been exposed to Agent Orange. He returned to his post at approximately noon with a physician's note identifying his condition as "symptoms compatible with irritant gas/fume exposure" and recommending transfer to the other section. On his return to the workplace, he described his diagnosis inaccurately as having been "petroleum distillate poisoning." Subsequently three other employees complained of an odor and were transferred to the east section. A building maintenance engineer could not identify an odor or exposure source.

At 2:00 p.m., supervisors of the company requested that employees report any odor or symptoms. Within a half hour, numerous new complaints of headache, nausea, and pharyngitis were received from employees. Supervisory personnel notified the city fire department, which evacuated the west section at 3:10 p.m. At 3:40 p.m., employees in the east section began to report similar symptoms. Ambulances were called, the entire building was closed, and fifty-eight employees were taken to seven local hospitals. Although nine employees were given intravenous fluids at the site, none was hospitalized after arrival to the hospital.

Industrial hygiene analyses were performed on the site at the time by the California Occupational Health and Safety Administration (CalOSHA), the city fire department, and the county Hazardous Materials Management Team. Finding no odors and no detectable air contaminants, CalOSHA permitted the building to be reopened at 6:45 p.m. on day 2. It was initially operated by supervisory personnel, with regular employees arriving that evening for the night shift.

On day 3 at 11:30 p.m., several employees again reported an odor and symptoms. The company asked the building to be evacuated immediately and ambulances were again called. At least one employee arriving for work was observed to report symptoms of respiratory distress before

entering the building and was given intravenous fluids by paramedics at the scene. Of the twenty-eight employees taken to the hospital, of whom five were among the initial group evacuated on day 1, none was hospitalized or given specific treatment.

Immediately following the second evacuation, and continuing through the evening of day 3, several air analyses were conducted in the building under various conditions of ventilation. By this time, the public agencies that had evacuated the building on day 1 were supplemented by corporate industrial hygiene and medical personnel as well as a private consulting firm with an industrial hygiene reference laboratory. No detectable air contaminants were found under normal conditions of ventilation, and only trace amounts of organic chemicals were detected after the ventilation system was shut down for several hours.

No structural or engineering changes were made and procedures were unchanged. One month later, a similar episode involving eleven employees occurred at about noon in the east section. The company's contract physician evaluated all but one of the affected employees on site and found no specific signs of illness. All but two employees returned to their workstation; the remaining two returned the next day. A repeat industrial hygiene evaluation was conducted by the industrial hygiene consultants at 12:40 p.m., with similar negative results. There have been no further complaints.

Each of the five agencies responding followed its own protocol for the evaluation of such incidents. The first to respond, on day 1, were three public agencies. The two teams dispatched by the company, a team of corporate personnel and a consulting firm on contract, arrived some hours later but were equipped to perform more exhaustive analyses during the second incident.

Initial evaluation by the public agencies emphasized a qualitative and semi-quantitative hazard evaluation. A walkthrough was conducted and numerous measurements were collected using direct-reading devices (Draeger tubes and hydrocarbon analyzers). Subsequently CalOSHA and the Hazardous Materials Management Team collected air samples using charcoal tubes, which were submitted to two laboratories for independent analysis by combined gas

**Table 11.8.** Psychogenic Outbreak Reconstruction: Industrial Hygiene Survey

Organic chemicals reaching detectable levels after ventilation system shut off for several hours (ppm):

	Result Obtained	Odor Threshold	OSHA PEL
Benzene	0.07	2.0	10
Hexane	4.5	220	500
Methylene chloride	1.5	200	500
Toluene	0.32	2.0	200
Trichloroethylene	0.1	20	100 (withdrawn)
Xylene	0.04	0.5	100
No detectable odor			

chromatography and mass spectrometry (GC-MS). Results of this investigation are provided in Table 11.8.

On day 2, corporate and consulting personnel and CalOSHA performed another walkthrough and sampling inspection, this time adding direct reading photo-ionization (H-Na), flame ionization, and electron capture gas chromatography to the initial analysis. They also repeated the collection of samples for GC-MS, including two-hour samples from air vents and working areas. These studies were performed during normal ventilation and after the ventilation system had been shut off for 24 hours.

A comprehensive evaluation of the ventilation system was performed using a velometer and smoke tubes, and by visual inspection of the inside of all accessible vents and ducts. The maintenance schedule of the ventilation system was reviewed, and odor tests were performed when the ventilation system was switched back on. All air filters (twenty-four in number) were replaced. The old filters were placed in one of the airtight bags, one for the intake and one for the exhaust system, and allowed to out-gas for a day. The atmosphere in the bags was then tested for volatile residues with a direct reading hydrocarbon analyzer.

An inventory of all cleaning compounds, solvents, degreasing agents, and equipment constituents was made by the company and

consultants. Air samples were taken from a janitor's closet and inside company vehicles. The manufacturers of the 125 video display terminals in use in the building were contacted to determine whether problems had been reported. Other potential sources that were evaluated by sampling, inspection, review of records, and review of constituents include room furnishings, drains and sewers, the street outside, insulation on the walls and in the ceiling, and the soil.

Except as noted, all industrial hygiene studies performed were negative. The presence of chemical contaminants could not be detected by any means until the ventilation system was shut down. Table 11.7 gives the chemicals identified, concentrations found, and pertinent reference information. No odor could be detected by individuals not complaining of symptoms, and only 68 percent of all employees reporting symptoms perceived an odor at the time. Of those that did, 83 percent were from stations in the east section, although only 41 percent of affected individuals in the first incident, 50 percent of those in the second, and 56 percent of those requiring intravenous fluids in the second incident were from that section. An odor was not apparent when the ventilation system was switched back on after 24 hours.

Following the industrial hygiene analysis of the incident, the emergency room records of affected individuals were obtained, and supervisory and selected employees were interviewed to reconstruct the sequence of events. With the cooperation and participation of the union representing the affected employees, an odor perception study was organized. In this study, a panel of blindfolded employees without a history of nasal disease smelled minute quantities of chemicals identified by the industrial hygiene consultants as present in the building in trace amounts, singly and in combination, to determine which, if any, most resembled the odor noted during the incidents. This exercise did not implicate a particular candidate chemical.

Medical reports of symptoms proved to be inconsistent. Table 11.9 gives the distribution of symptoms affecting more than one-third of the presenting employees by the health facility to which they presented. Reported symptoms varied considerably by health facility. Table 11.10 presents the clinical findings reported by the treating

**Table 11.9.** Psychogenic Outbreak Reconstruction: Symptoms by Building Occupants

Facility	Number	Symptom			
		Headache	Nausea	Dizziness	Pharyngitis
Clinic	19	X			X
Hospital A	17	X		X	
Hospital B	8	X	X		
Hospital C	6				X
Hospital D	5	X		X	
Hospital E	5	X	X		
Hospital F	5		X	X	
Hospital G	4			X	X
Hospital H	3				X
Hospital I	73	X	X		

**Table 11.10.** Psychogenic Outbreak Reconstruction: Clinical Findings Reported by Treating Physicians

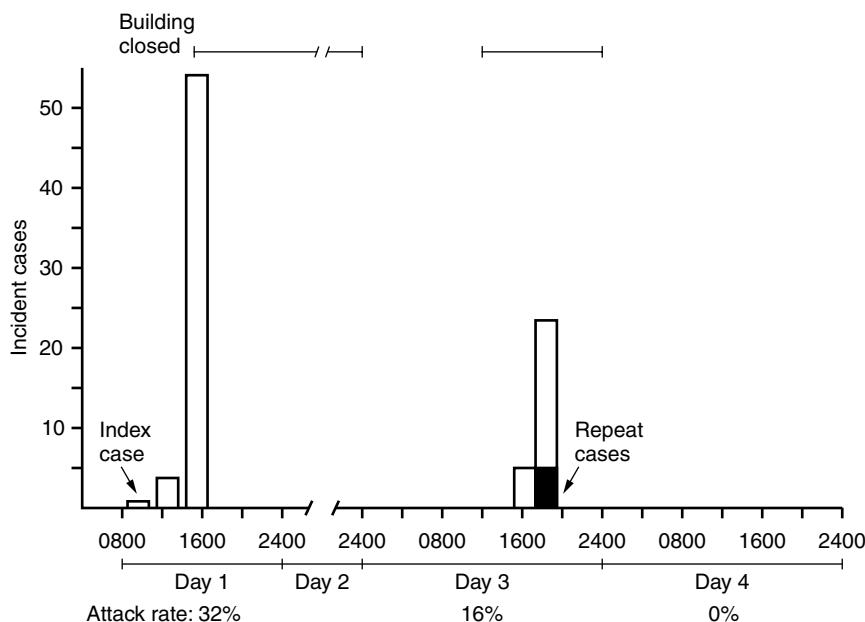
- Non-specific symptoms, inconsistently reported
  - Headache
  - Nausea
  - Dizziness
  - Pharyngitis
  - Weakness
  - Hyperventilation (8/15 cases)
- Diagnostic terms used
  - “Irritant gas/fume exposure”
  - “Allergy to fumes”
  - “Chemical inhalation”
  - “Status post fume exposure”
  - “Pregnant”
- Case(s) reported ill without exposure (sent to hospital from parking lot)

physicians; these symptoms were also inconsistently reported. All symptoms abated rapidly without specific treatment, and no patients were hospitalized. At Hospital A, fifteen of seventeen employees seen had arterial blood drawn for blood gas analysis; eight of the fifteen showed hypocapnia, diagnostic of hyperventilation.

The epidemic curve that characterized the two days of the first incident is presented in Figure 11.1. A reconstruction of the sequence of employee transfers and illness is given in Figure 11.2, on the model of the propagation of a transmissible agent.

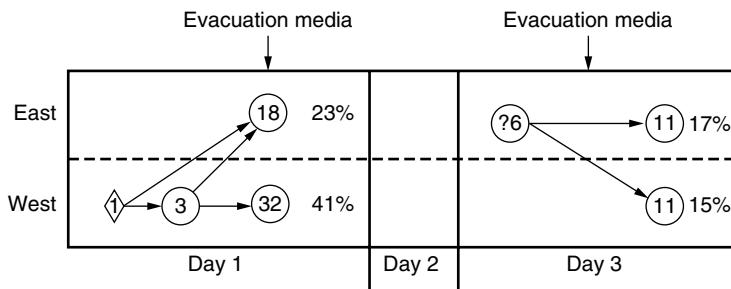
No other reports of illness or odor were received by occupants of other buildings in the area, and no spill or unusual traffic congestion was apparent on the nearby freeway or in the city streets. The sewer system was not backed up and hydrogen sulfide was not detected in the building.

INCIDENT RECONSTRUCTION: EPIDEMIC CURVE



**Figure 11.1.** Epidemic curve for a building-associated outbreak of psychogenic origin. Adapted from Guidotti TL, Alexander RW, Fedoruk MJ. Epidemiologic features that may distinguish between building-associated illness outbreaks due to chemical exposure or psychogenic origin. *Journal of Occupational Medicine* 1987; 29: 148–150. Reprinted with permission.

## INCIDENT RECONSTRUCTION: PROPAGATION TREE



**Figure 11.2.** Propagation trees for a building-associated outbreak of psychogenic origin. Adapted from Guidotti TL, Alexander RW, Fedoruk MJ. Epidemiologic features that may distinguish between building-associated illness outbreaks due to chemical exposure or psychogenic origin. *Journal of Occupational Medicine* 1987; 29: 148–150. Reprinted with permission.

Interviews with affected employees confirmed that the index case had actively spread alarming reports of toxic exposure during each outbreak. This index case was also reported by several witnesses to have claimed that he had tested the accuracy of the industrial hygiene analysis by soaking his shoes with gasoline. It was then determined that the index case had received military training in psychological warfare prior to his present employment. He was dismissed the following day.

This incident is unusual both in the degree of thoroughness with which the building was studied as well as the degree to which the early events in the incident can be established, a reflection of this company's remarkable occupational health capabilities.

This incident is as clearly negative for physical or chemical exposures as any case yet reported. Environmental analysis was conducted on a timely basis by the most current methods by five different groups using different procedures and equipment. The results are consistent in ruling out a toxic exposure. Furthermore, a re-challenge of the same population (including most of the same individuals) under the same conditions did not result in another outbreak. When a third incident did occur weeks later, it was aborted by reassurance alone, and no documented illnesses resulted. Thus, this incident has ruled out an exposure within the building with at least as much and

probably more certainty than a negative challenge test in testing for allergy or an adverse drug reaction.

This case stands as an unusually clear example of a psychogenic building-associated outbreak. It established the propagation pattern as a positive indicator of psychogenic outbreaks and demonstrates well the cardinal features of such incidents.

## INTERVENTIONS

Building operators may belong to the Building Owners and Managers Association (BOMA), an international organization based in Washington, D.C. ([www.boma.org](http://www.boma.org)). BOMA provides information and training programs within the industry and is a source of expertise that building operators trust.

Building-associated outbreaks that are associated with a definite chemical or physical cause can be managed appropriately for the problem. Sources of exposure to infectious agents or antigens can be identified and removed. Sources of chemical exposure or odor can be removed. Areas of dampness can be corrected. Emissions within the building, such as those from kitchens, may be vented to the outside through a local exhaust system. The bigger management problem lies with those cases in which a specific exposure cannot be identified.

Buildings for which there is a possibility of out-gassing from new furnishings, carpet, or construction materials can be “baked.” This involves increasing the temperature as much as management feels comfortable with and then ventilating the building at maximum capacity, usually over a weekend. This increases the rate of evaporation, sublimation, and out-gassing of volatile constituents and exhausts them from the building. The temperature should be turned back down well before workers return to the building.

Interventions for the management of SBS associated with poor indoor air quality are limited but are normally adequate.

The ventilation rate for the building can be increased. In an important 1993 study of reported symptoms and perception by Menzies et al., increased air exchange was shown to have no

demonstrable effect on reported symptoms. However, the study did not report a detailed evaluation of the buildings in which the trial took place. Ventilation is likely to be most effective when there is evidence of poor indoor air quality and when occupants of the building are aware of the change.

Ducts can be cleaned, and routine maintenance can improve the functioning of a neglected HVAC system. Filters should be changed at least as often as recommended by the manufacturer.

For individual cases of building-related illness, local desktop air-cleaning units to remove dust are a possibility, but they must be placed close to the breathing zone of the affected individual. These units are not effective beyond a foot or so. Electrostatic units are to be avoided because they generate ozone and may make the problem worse. Relocation of the affected worker to another office may be effective. Management is often tolerant of one such move but will usually not tolerate a second request or a request to work from home, and consequently may become uncooperative. The affected worker needs to know this before making such a request. Area cleaning is important, but if the affected worker does it himself or herself, there may be additional exposure to dust or cleaning agents. On the other hand, if it is done by custodial staff, the quality of the cleaning cannot be assured.

Psychogenic outbreaks are extremely difficult to manage because the explanation will inevitably be rejected. The literature recommends education and communication, but in the course of such an event it is almost impossible to be heard, even if one is conveying accurate information. The principles of risk communication should be applied (see Chapter 8), but the most effective approach is probably for management to correct, and to be seen to be correcting, some of the underlying issues in the workplace at the same time that the hygiene investigation is in process. Affected workers will expect and may demand that the hygiene survey be comprehensive. Once that is achieved and negative findings are communicated back to the workers, other explanations may be possible, but not everyone will be convinced. Some people will always believe that a mysterious substance X was overlooked, or that knowledge of it was suppressed.

## RESOURCES

- Burroughs B, Hansen SJ. *Managing Indoor Air Quality*. 4th ed. Lilburn GA: Fairmont Press; 2008. [Excellent guide for building owners and managers, useful to physicians only if this is a major part of one's practice.]
- Rogers R, Guidotti TL. Indoor air quality and building-associated outbreaks. In: Cordasco E, Demeter S, Zenz C, eds. *Environmental Lung Disease*. New York: Van Nostrand Reinhold; 1995:179–207.
- Spengler JD, McCarthy JE, Samet JM. *Indoor Air Quality Handbook*. New York: McGraw-Hill; 2000. [Comprehensive but very expensive and dated. Unfortunately, there are few good books that are both current and suitable for physicians on this topic.]

## NOTEWORTHY READINGS

- Alexander RW, Fedoruk MJ. Epidemic psychogenic illness in a telephone operators' building. *J Occup Med*. 1986;28:42–45.
- Bauman MGD. Volatile organic compound emissions from composite wood products: a review. Madison, WI: US Forest Service, Forest Products Laboratory; 1997. Available online at <http://www.fpl.fs.fed.us/documents/pdf1997/bauma97a.pdf>. Accessed November 7, 2008.
- Cole TB, Chorba TL, Horan JM. Patterns of transmission of epidemic hysteria in a school. *Epidemiology*. 1990;1:212–218.
- Cone JE, Hodgson MJ, eds. Problem buildings: building-associated illness and the sick building syndrome. *Occup Med*. 1989;4(4):575–802.
- Gabbage RD, Kaye SV. *Indoor Air and Human Health*. Chelsea MI: Lewis; 1985.
- Gann PH. Truth or consequences: when the study reintroduces the disease agent. *Epidemiology*. 1990;1:192–194.
- Gottlieb NH et al. Impact of a restrictive work site smoking policy on smoking behavior, attitudes, and rooms. *J Occup Med*. 1990;32:16–23.
- Guidotti TL, Alexander RW, Fedoruk MJ. Epidemiologic features that may distinguish between building-associated illness outbreaks due to chemical exposure or psychogenic origin. *J Occup Med*. 1987;29:148–150.
- Kreiss K. The sick building syndrome: where is the epidemiologic basis? *Am J Public Health*. 1990;80:1172–1173.
- Kreiss K, Hodgson MJ. Building-associated epidemics. In Walsh PJ, Dudney CS, Copenhaver ED, eds. *Indoor Air Quality*. Boca Raton FL: CRC Press; 1984: 88–103.
- Macher JM, Huang F-Y, Flores M. A two-year study of microbiological indoor air quality in a new apartment. *Arch Environ Health*. 1991;46:25–29.

- Menzies D, Bourbeau J. Building-related illness. *N Engl J Med.* 1997;337(21):1524–1531.
- Menzies R, Tamblyn R, Farant J-P, Hanley J, Nunes F, Tamblyn R. The effect of varying levels of outdoor-air supply on the symptoms of sick building syndrome. *N Engl J Med.* 1993;328(12):821–827.
- Norback D, Michel I, Widstrom J. Indoor air quality and personal factors related to the sick building syndrome. *Scand J Work Environ Health.* 1990;16:121–128.
- Ooi PL, Goh KT. Sick building syndrome: an emerging stress-related disorder? *Int J Epidemiol.* 1997;26(6):1243–1249.
- Salthammer T. Emissions of volatile organic compounds from furniture coatings. *Indoor Air.* 2004;7(4):189–197.
- Spitzer WO et al. Links between passive smoking and disease: a best-evidence synthesis. A Report of the Working Group on Passive Smoking. *Clin Invest Med.* 1990;13:17–46.
- Willes SR, Fitzgerald TK, Bascom R. Nasal inhalation challenge studies with sidestream tobacco smoke. *Arch Environ Health.* 1992;47:223–230.

*This page intentionally left blank*

# **12 ENVIRONMENTAL HEALTH**

Environmental medicine is the “other half” of occupational and environmental medicine (OEM). The complementary relationship between environmental medicine and occupational medicine is explored in Chapter 1, with notes on how environmental medicine may be practiced. The intellectual terrain is mapped in this chapter, with notes on the role of physicians in the protection of environmental integrity. This chapter is titled “environmental health” rather than “environmental medicine” because it outlines environmental issues from a public health perspective, which is more applicable to the protection of populations and the community than is a clinical, patient-by-patient approach. It is not a comprehensive introduction to environmental health. Rather, it emphasizes the environmental health issues most likely to be encountered in the course of the practice of occupational and environmental medicine, and addresses some issues that are not often discussed in the environmental health literature.

Environmental hazards are mostly the same as occupational hazards. The centrality of exposure assessment in understanding the significance of exposure is the same. The technology for measurement of exposure varies somewhat with the scale of the measurement, but many methods are the same, and the interpretation of the findings is

similar. The physiological principles are the same. The principles of toxicology and epidemiology are identical.

Environmental hazards and occupational hazards differ primarily in context, in degree of exposure in various settings, and in the population exposed. Environmental exposures act subtly over large populations, which include many individuals who are not healthy, as well as many who may be biologically susceptible (the very young, the very old, women and the fetus during pregnancy, and the chronically ill). These large populations also include those who may be vulnerable, such as people who live in a particular place or who cannot protect themselves due to income, class, past discrimination, and justice issues not related to workplace justice. In contrast to occupational disorders, individual cases of documented environmental disease are uncommon. Environmental exposures are much more likely to be risk factors and determinants in population health than the drivers of disease risk. In other words, they act on a population level to influence the distribution of disease (such as asthma triggered by air pollution), but not as the discrete or exclusive cause of a particular individual's disease. (For example, there is no "air pollution disease," just an increased risk of bad outcomes in a variety of diseases associated with air pollution, among other risk factors.)

Historically, occupational physicians have tended to be called in on problems involving air quality and hazardous waste, where the chemical regimen is similar to the workplace. General preventive medicine and (medical) public health specialists have been called in on issues involving water and food safety, because their core expertise is public health protection and thus more often involves infectious disease. These historical tendencies are changing with the consolidation of responsibilities among consultants and medical directors, increasing litigation, concern over consumer protection and liability issues, globalization and responsibility for foreign operations, improved training of occupational physicians in public health generally, and the threat of emerging infections in the workplace.

The OEM physician deals with a different range of environmental issues than the public health or preventive medicine specialist,

although there is much overlap. Table 12.1 lists specialty areas of environmental health, supplemented by topics named as core on the specialty certification examination for occupational medicine by the American Board of Preventive Medicine. It is clear that the OEM physician does not need to master everything on this list. Environmental health services are normally provided by local public health departments. The OEM physician needs to know about some of them at a level of understanding sufficient to allow consultation with

**Table 12.1.** Traditional Public Health Functions in Environmental Health

- 
1. Air quality management (monitoring and control of air pollution)
  2. Consumer safety
  3. Control technology (engineering and control technology)
  4. Emergency preparedness (disaster management)
  5. Environmentally related chronic disease epidemiology
  6. Environmental medicine
  7. Food protection (toxicology and microbiology)
  8. Genetics (including epidemiologic studies and studies of genotoxicity)
  9. Housing conservation and rehabilitation
  10. Institutional environmental health (sanitation in schools, health facilities, prisons, etc.)
  11. Noise control
  12. Radiation health physics
  13. Recreational health and safety (control of injuries and health problems associated with sports and recreation)
  14. Risk assessment
  15. Solid waste management
  16. Toxic substances management (including hazardous waste control)
  17. Toxicology (study of health effects, handling by the body, and pertinent factors of the host influencing the response to toxic agents)
  18. Traffic safety
  19. Vector control
  20. Water supply and treatment
  21. Water quality management (including wastewater treatment)
-

responsible authorities when a problem emerges within his or her scope of responsibility and to initiate a reasonable response, with more thorough preparation to follow.

In practice, OEM physicians must thoroughly master the workplace environment and the relationship between releases inside and beyond the plant boundary. They need a rather broad knowledge of air quality issues, of the toxicology of common environmental pollutants (most of which are the same chemicals found in the workplace), and of common infectious hazards with an emphasis on zoonoses (a major driver of emerging infections in the modern world). In addition, they must understand public health protection as practiced in facilities management (food, water, sanitation), and life-cycle issues of hazardous chemicals of importance in industry, including waste. Obviously, additional, highly specific knowledge is required for anyone working in an environmentally sensitive industry, such as oil and gas.

The OEM physician working within a large organization, such as a corporation, will most often be involved in decisions and guidance to management, on issues involving human health risk assessment, regulatory compliance, product stewardship, and environmental management at the employer's facilities. In practice, this means that OEM physicians with the responsibility for environmental issues within their organizations need a working knowledge of "traditional" environmental health and a good working knowledge of ecosystem issues, with special reference to any issues related to sensitive products or services. They must also have a broad background in the regulatory issues related to products and services (often involving regulations in the United States, Canada, the European Union, and elsewhere) and a familiarity with environmental health in the built environment, including building-associated illness. The more the OEM physician knows, the more effective he or she will be, especially when a problem is newly emerging. However, nobody expects the OEM physician to be an environmental engineer or environmental public health officer.

In the corporate sphere, standards for environmental management are defined by the International Organization for Standardization

(ISO, which is not an acronym but a clever play on the initials: *iso* in Greek means equal, implying standardization). ISO is a highly authoritative and respected organization based in Geneva, the membership of which consists of major corporations and government agencies. It develops voluntary standards for industry through a network of technical committees. ISO 14000 is the series that pertains to environmental management and defines a process characterized by policy, planning, supervised implementation, monitoring results, review, and improvement. ISO 14000 does not prescribe exposure or ambient environmental standards; other bodies already do that, and government regulatory agencies have the authority of law.

## AIR QUALITY

Simply put, air pollution is the emission into the air of chemically active compounds at a rate that exceeds the capacity of natural processes to convert, transport, dilute, or dissipate them. Air pollution affects health when these compounds accumulate to concentrations capable of producing a biologically significant effect. Until recently, these effects were exceedingly difficult to measure accurately, but advances in environmental epidemiology have demonstrated that air quality is a major determinant of human health in cities (see Chapter 3).

The accumulation of chemically active compounds in the atmosphere is greatly affected by land features and by atmospheric movements. Valleys, close-in mountain ranges, and coastal plains tend to confine and concentrate air pollution. These features hold the air mass like a container, and prevent dilution and mixing. Stagnant air masses may receive emissions for days on end. When winds are calm, a thermal layer of warmer air may sit on an air mass like a lid on a jar and prevent its movement. This is called an “inversion” because the natural tendency for air temperature to fall with altitude is reversed. Inversions are common but tend to be frequent and long-lasting in only a few areas, principally in narrow river valleys and, notably, along the coast of southern California. Lack of open space (parks, forests, wilderness areas, bodies of water) also

strongly affects air pollution in a locale by reducing dilution and the mitigating absorbing and bio-transforming effects of plants. The combination of high population density, reliance on the automobile, heavy emissions, enclosure by mountains, and frequent inversions explain why air pollution has been a particular problem in Los Angeles.

### **Types of Air Pollution**

Table 12.2 presents a typology of community-level, ambient air pollution, defined by different chemical characteristics, distribution, and sources. However, features of each type may be present to varying degrees in any given community's air pollution. Indoor air quality is discussed in Chapter 14. Atmospheric change is discussed at the end of this chapter in the section on ecosystem health.

“Reducing” air pollution is caused by the emission of sulfur dioxide ( $\text{SO}_2$ ) and particles that are chemical reducing agents in the atmosphere. (See Chapter 10 for discussion.) This is by far the oldest type of air pollution, and was responsible for the infamous London “killer fog” of the 1950s, which resulted in thousands of deaths, and the repeat of 1962, which resulted in hundreds of deaths. It was also responsible for several air pollution incidents in the early and middle twentieth century, most notably in the Meuse Valley of Belgium in 1930 and a lethal episode in Donora, Pennsylvania, in 1948. Emissions of  $\text{SO}_2$  are caused by burning fossil fuels containing some sulfur, producing particles across all size ranges. Emissions of particles occur most heavily when combustion is inefficient and much of the fuel goes unburned. Reducing air pollution was historically produced primarily by stationary combustion sources, such as fossil fuel power plants, industrial furnaces, and home heating units. This type of air pollution predominated in older concentrations of basic industry in the Northwest and Midwest. It remains a problem in many parts of the developing world, and until relatively recently was still the major problem

**Table 12.2.** Types of Air Pollution

Type	Composition	Source
Reducing	Sulfur dioxide, particulates	Stationary combustion sources, such as fossil fuel power plants, industrial furnaces, home heating units
Photochemical	Hydrocarbons and oxides of nitrogen, ozone (secondary pollutant), aldehydes, organic nitrates, fine particulate matter	Mobile emission sources, internal combustion engines (diesel and gasoline), point sources of reactive or catalytic air toxics
Point-source air toxics	Specific chemicals, not regulated by NAAQS	Industrial and commercial, sometimes accidents
Indoor	See Chapter 11	Home heating, cooking, combustion sources in the building, passive cigarette smoke
Long-range	Sulfates, oxides of nitrogen, mercury (power plants)	Stationary sources
Global atmospheric change	Carbon dioxide, methane, water vapor, ozone-depleting chemicals (many of which are also associated with greenhouse effect)	Most human activity, including industry and agriculture

in China, although it has been overtaken by photochemical air pollution. Table 12.3 compares the characteristics of reducing and oxidizing, or photochemical, air pollution. Reducing air pollution is now mainly of historical interest in North America, but remains a problem in many cities in developing countries.

**Table 12.3.** Air Pollution in U.S. Metropolitan Areas

	Oxidizing	Reducing
Chemistry	Photochemical, free radical, fine particulate matter	Reducing, acid hydration
Predominant emissions	$\text{NO}_x$ , particulate matter	$\text{SO}_x$ , particulate matter
Emissions source	Automotive exhausts, diesel, stationary sources	Stationary source, esp. coal combustion
Geographical distribution in North America	Originated in Sun Belt of United States, now ubiquitous	Mostly older industrial communities of Northeast and North Central, now effectively controlled
Primary control strategy	Catalytic converter for automobiles, source controls	Retrofit of stationary sources
Meteorological conditions	Sunlight, atmospheric inversions	Fog, atmospheric inversions
Major event	Pasadena, 1943; subsequent events in 1950s	Donora, 1948

Photochemical air pollution is much newer in human history; it was initially an urban phenomenon but is now global. Together with fine particulate air pollution, it is now the predominant type of air pollution worldwide. Photochemical air pollution results from complicated chemical reactions in the atmosphere that are driven by the energy in sunlight. In photochemical smog, emissions rich in oxides of nitrogen and hydrocarbons undergo reactions driven by sunlight to produce ozone, specific compounds of nitrogen, and aldehydes, all of which are highly reactive and chemically oxidizing. This type of smog is caused primarily by emissions from mobile sources, mostly from automobiles. However, it is also exacerbated by certain types of

emissions from stationary sources, such as evaporated hydrocarbons from gasoline and dry cleaning solvents, which catalyze the photochemical reactions, and oxides of nitrogen from power plants, which serve as substrates. Because of the need for sunlight to drive these chemical reactions and because of local reliance on the automobile for transportation, photochemical air pollution appeared first in the cities of the sun belt, especially Los Angeles. The increasing proliferation of diesel-powered trucks using older diesel technology added increasing amounts of fine particulate matter (particles 2.5  $\mu\text{m}$  or less in aerodynamic diameter, which is discussed in Chapter 2) to that already present from combustion sources (but unmeasured in that era).

A third type of community air pollution is point-source emissions, often associated with the release of “air toxics.” These pollutants are not a consistent feature of ambient air pollution but have potential health or ecological effects. Examples include perchlorethylene or “perc” (which is released from dry-cleaning plants), methylene chloride (a solvent), and asbestos. Air toxics that are released from point sources may become a local problem affecting the immediate vicinity of the plant, or they may be more widely distributed and contribute to air quality in an entire region, as in the case of benzene from evaporated gasoline. Air toxics are not regulated under usual ambient air quality standards in the United States.

Indoor air pollution is discussed in detail in Chapter 11. Attempts to achieve energy efficiency by insulating buildings tightly contributed to this problem. Many substances that play only a minor or negligible role in outdoor air pollution are much more important indoors, such as formaldehyde, asbestos, and sidestream cigarette smoke.

Acid deposition (or “acid precipitation,” often called “acid rain”) is a general term for the long-range transport of acidifying chemical substances, principally sulfates but also oxides of nitrogen. Precipitation brings the acidifying chemicals to the ground and into lakes. The result is a change in pH that had a devastating effect on aquatic life and terrestrial plants. The principal disturbances caused by acid

deposition appear to be ecological, with acidification of lakes and effects on sensitive species of trees. However, some studies suggest human health effects in the form of increased frequency and severity of asthmatic attacks. This problem was exacerbated by efforts to control ground-level pollution by building higher stacks; the pollutants then rose much higher and were transported much longer distances. Acid deposition became an international problem, as emissions from the Midwest traveled freely over Canada, and emissions from Germany and Central Europe fell on Scandinavia.

Today the problem has been greatly reduced by much more stringent emissions controls. Thanks to pioneering work done by aquatic biologists who documented the effects on the scale of entire lakes, governments were persuaded to impose more stringent regulations on acidifying emissions, and the problem was largely reversed. Although it has received little attention from the public, the resolution of this problem is considered within air quality management circles as one of the great environmental success stories of the twentieth century. However, by comparison to other environmental problems, it was a relatively simple one.

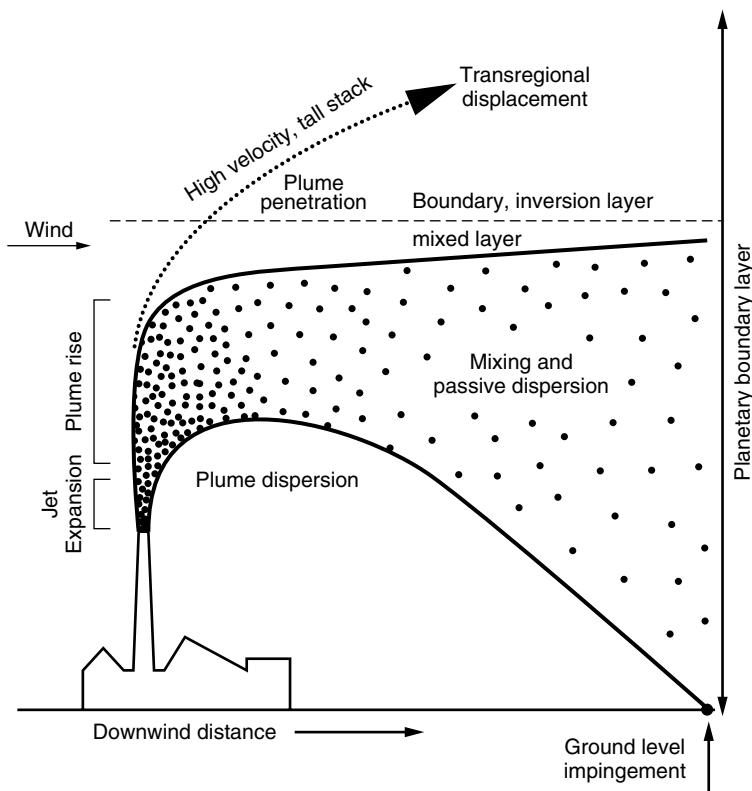
### **Air Pollution Dispersion**

Air pollution travels with wind, and concentrations dilute with distance and mixing. Models that quantify this process are called “air pollution dispersion models” and are frequently used to anticipate the impact on air pollution of new facilities, proposed regulations, modifications to plants, and incidents involving the release of air contaminants. Wind speed, direction, and flow rate at the source are critical variables for the models. Meteorological conditions, obstacles to flow (such as tall buildings or mountains), and the terrain that the wind passes over all have major effects on the predictions of such models. Models that are more sophisticated can calculate dispersion patterns for many overlapping sources as well as for highly dispersed sources, and can even track pollution back to the source, given patterns in their concentration downstream.

An “inversion layer” is a key meteorological phenomenon that affects the dispersion of air pollution. Ordinarily, the ground is warmer than the air during the day, and therefore the air just above it is warmer than the air higher up; temperature falls with altitude. However, there are many situations in which this is not the case, and air temperature is higher at higher altitudes. Because this is the reverse of the usual situation, it is given the name “temperature inversion.” Inversions often happen early in the morning, before the ground heats up. At this time, the air will be coldest next to the ground, and the temperature will increase with altitude up to a point (which is the reverse of the normal case), and then after that point will begin to cool again. Inversions also occur when warm masses of air come off the ocean, or when a warm front sits on top of a cooler air mass—usually because both are trapped in a valley or natural basin, such as Los Angeles. Inversions reduce the mixing between air masses and act like a lid on a jar, restricting air movement, because the cooler air underneath, being denser, cannot rise into the less dense air mass above it. The level at which this capping effect occurs is called the “boundary layer” or “inversion layer.” Inversion layers are broken up when the air mass becomes unstable because of wind or intrusion of warm fronts.

The most basic dispersion situation is illustrated in Figure 12.1, depicting a cool, clear day with a slow, steady wind blowing toward the right in the figure. Air conditions are stable and there is an inversion layer as indicated. A tall stack is receiving emissions from a facility, which may be a factory, a power plant, or any other stationary source. The air, in general, is warmer than the atmosphere and exits the stack with a certain velocity. If the velocity of the exiting gas is high and the temperature exceeds that of the surrounding air, there is a “jet” effect as the gas enters the cooler atmosphere and continues to rise, buoyant, expanding as it leaves the stack. If the stack is tall enough, the emissions can go quite high and even through the boundary, where they may be carried long distances by air currents. More often, the “plume” of gas exiting the stack reaches more or less neutral buoyancy and is constrained by the boundary layer so that it

cannot expand upward. Tilted by the steady wind, it forms a rough, lopsided cone, diluting as it expands, until the plume eventually touches down to earth at a point some distance (often many miles) from the stack. This point is called the “fumigation point” and constitutes the maximum concentration of emissions at ground level due to the plume. The taller the stack, the farther away the fumigation point. If there is no inversion, the discernible plume expands indefinitely and is eventually lost in the atmosphere. If there is a stronger wind, there is likely to be more mixing and much more rapid dilution, so the concentration of emissions will fall much more rapidly with distance from the stack. If there is turbulence, mixing may be so



**Figure 12.1.** A simple model of plume dispersion from a point source of air pollution.

complete in the air immediately downwind from the source that no plume can be discerned.

### **Health Effects Associated with Air Pollution**

The toxicology of air pollution is exceedingly complex. There are different types of air pollution, as noted; many different pollutants; and great variation in individual susceptibility to their effects at low concentrations.

Air pollution has been associated in several severe episodes with high mortality, usually among persons with pulmonary or cardiovascular disorders. While mostly of historical interest today, in part because of emissions controls and in part because the type of pollution (the reducing variety) is less common today, these incidents still serve as warnings on many levels. Although a series of more serious air pollution episodes preceded the Second World War, a particularly intense episode occurred in Los Angeles on September 8, 1943, which led to considerable public outrage. It provided the impetus for sustained political pressure, led by the city of Pasadena, on the County Board of Supervisors and the California state legislature. This resulted in the formation of the California Air Resources Board and the Air Pollution Study Project, which had to invent its own methodology as it went along. At first it was assumed that air pollution in California was like that in the East, resulting from sulfur oxides. A series of ineffective measures eroded public confidence in the ability of the government to deal with the growing problem. Valuable time was lost before Arie J. Haagen-Smit, of the California Institute of Technology, identified the phenomenon of photochemistry in 1952. John R. Goldsmith then developed epidemiological methods for documenting health effects, and air quality management was put on a more productive path. Investigators were also misled at first because they thought that the same factors were operating that had caused the 1948 Donora incident.

Table 12.4 presents a concise summary of major health effects thought to be caused or exacerbated by community air pollution. This continues to be a very active area of research, and new findings appear frequently.

Mortality is primarily associated with fine particulate air pollution and ozone, and appears to show no threshold. Although principally due to cardiovascular causes, mortality has increased in a manner not simply due to “harvesting” (the technical term) premature deaths of

**Table 12.4.** Diseases or Conditions for Which Air Pollution Exposure May be a Contributing Factor

Condition	Association with Air Pollution	Associated Factors	Diagnostic Approach
Eye irritation	Aldehydes, especially PAN	Atopy, low threshold for mucosal irritation	History
Bronchitis	Direct irritation	Smoking, pre-existing allergies or atopy	History. Consider aggravation of allergic bronchitis or bronchitis due to smoking
Lung cancer	Particulate matter	Smoking	Statistical correlation; cannot diagnose in individual
Asthma and aggravation of reactive airways	Ozone, PM, NO <sub>x</sub> , trigger for reactive airways, inflammation	Atopy, pre-existing asthma and allergies	Correlation with ozone, particulate matter, other air pollution indices
Headache	Carbon monoxide	Smoking	History of exposure and elevated CO levels
Lead toxicity	Major source of exposure until 1980s	Lead removed from gasoline in North America	Blood lead and thorough exposure history

ill people who would have died anyway. The source of this increase in mortality took a very long time to discover and document. It was particularly difficult to identify the component of air pollution primarily responsible for this increased mortality because pollutants in modern photochemical air pollution “covary” (move together) as well as varying with the weather.

New methods emphasizing time series analysis (see Chapter 3) provided a methodological breakthrough in the 1980s that conclusively demonstrated high levels of excess deaths associated primarily with fine particulate air pollution. This became a public health priority in the 1990s because the initial studies on air pollution had used relatively conventional, epidemiological study designs emphasizing relative risk, which tended to minimize the effect against the normal mortality rate for a large city. When the risk estimate was changed to attributable risk, it became obvious that thousands of avoidable deaths were occurring in direct response to air pollution episodes, once weather (hot and humid conditions) and other confounding factors were taken into account. (See Chapter 3.)

Respiratory symptoms are the most common effects of air pollution of all types. Patients with asthma and chronic obstructive pulmonary disease (COPD) often experience a worsening of their symptoms during air pollution episodes, particularly when ozone levels are high. Whether ozone is a cause of asthma is still unclear, but its role as a trigger is well established. Individuals who do not have clinically diagnosed asthma but who have reactive airways or a history of atopy may experience a cough due to increased irritation of the bronchial mucosa. Common symptoms experienced even by people without reactive airway disease during air pollution incidents include cough, nose and throat irritation, and mild shortness of breath, which is an effect of ozone on the lung. These respiratory symptoms are often associated with eye irritation and a sense of malaise. Athletes often report that their performance is off and that they tire more rapidly when exercising during periods of high pollution levels. Acute upper and lower respiratory tract infections appear to be more frequent in residents of areas with higher pollution levels.

Cardiovascular effects of air pollution are associated primarily with fine particulate levels, as a risk factor for cardiovascular mortality in populations, and carbon monoxide, as a risk factor for individuals, because it both reduces oxygen delivery to the myocardium acutely and accelerates the process of atherosclerosis. These effects may occur in normal individuals without unusual susceptibility. Respiratory effects of air pollution, particularly complicating chronic bronchitis, may place an additional strain on cardiac function.

Mucosal irritation in the form of acute or chronic bronchitis, nasal tickle, or conjunctivitis is characteristic of high levels of air pollution, although individuals vary considerably in their susceptibility to such effects. Frequently the eye irritation is particularly severe, usually in a setting of high levels of particulates (which need to be in the respirable range described, and may be quite large soot particles) or of high concentrations of photochemical oxidants and aldehydes. This can make tasks requiring visual acuity and concentration difficult and can cause problems for those wearing contact lenses.

Cancer has always been a concern, but until recently there was little evidence to suggest that community air pollution was a significant cause of cancer. However, recent studies have used the more advanced methodology of environmental epidemiology, with better exposure assessment and based on much larger populations. These studies suggest that lung cancer rates are indeed associated with fine particulate air pollution, and the association is now accepted.

Neurobehavioral effects in children may result from accumulated body burdens of lead, to which air pollution contributed a large fraction prior to removal of lead from gasoline in the 1980s. Although leaded gasoline is now banned in North America, it remains a hazard in some developing countries, and its legacy remains in lead-contaminated soil and house dust, particularly near highways and in cities.

These health effects are better characterized for populations than for individual patients. Establishing a relationship between the symptoms of a particular patient and exposure to air pollution is much more difficult than interpreting the likely health effects on an entire community.

Although exposure to air pollution is inevitable in urban areas, there are specific measures physicians can recommend to protect patients who may be susceptible to its effects. Children, the elderly, and the chronically ill should be protected from excessive exposure to air pollution. During peak air pollution periods, exercise outdoors should be kept to a minimum, because the adverse effects of CO total oxidants may be increased during periods of exertion, and increased minute ventilation greatly increases exposure of the lower respiratory tract to particulates and penetrating gaseous pollutants. Appropriate ventilation in the home will reduce exposure to indoor air pollution, although at the cost of energy efficiency. Air conditioning is somewhat effective in cleansing the atmosphere of pollutants, and it does help and reduces the pollen count, which is important for persons with asthma. Air conditioning can partially protect the elderly and ill from the synergistic risk of high temperature and air pollution. Although it is desirable that children be raised away from heavily polluted neighborhoods, relocation of the family to a new home may be financially impossible, as well as socially and psychologically disruptive. Advising such a move when it is unrealistic for the family may only serve to increase their anxiety and produce frustration.

Patients with asthma, chronic bronchitis, and emphysema should be warned that they may experience exacerbations during periods of heavy air pollution. For some, supplemental bronchodilators, anti-inflammatory treatment, or restriction of activity may be necessary at times.

### **Regulation of Air Pollution**

Air pollution in the United States is ultimately regulated by the Environmental Protection Agency under the Clean Air Act of 1970 and its amendments of 1990. The Clean Air Act has been a cornerstone of the regulatory framework for environmental protection from the beginning of the agency's history.

The EPA initially recognized six pollutant categories as being widespread and interrelated problems in air quality: particulate matter, oxides of sulfur ( $\text{SO}_x$ ), oxides of nitrogen ( $\text{NO}_x$ ), ozone ( $\text{O}_3$ ),

carbon monoxide (CO), and lead (Pb). Particulate matter is now divided into PM<sub>10</sub> (“coarse particulate matter,” 10 µg and under in aerodynamic diameter) and PM<sub>2.5</sub> (“fine particulate matter,” 2.5 µg and under). (These pollutants are discussed individually in Chapter 10, and the behavior of particles is presented in Chapter 2.) The seven pollutant categories are regulated under the National Ambient Air Quality Standards (NAAQS), which are described in Table 12.5.

**Table 12.5.** National Ambient Air Quality Standards

Pollutant	Primary NAAQS	Averaging Times	Secondary
Carbon monoxide (CO)	9 ppm = 10 mg/m <sup>3</sup> 35 ppm	8 h (once/y) 1 h peak (once/y)	None
Lead (Pb)	0.15 µg/m <sup>3</sup> (2008)	Quarterly average	None
Nitrogen dioxide (NO <sub>2</sub> )	0.053 ppm = 100 µg/m <sup>3</sup>	Annual arithmetic mean*	= Primary
Particulate matter			
PM <sub>10</sub>	50 µg/m <sup>3</sup>	Annual arithmetic mean* (3 y ave.)	= Primary
	150 µg/m <sup>3</sup>	24 h (once/y)	
PM <sub>2.5</sub>	15.0 µg/m <sup>3</sup>	Annual arithmetic mean* (3 y ave.)	= Primary
	65 µg/m <sup>3</sup>	24 h (3 y ave. of 98th percentile)	
Ozone (O <sub>3</sub> )	0.075 ppm (2008)	8 h (3 y ave of 4th highest)	= Primary
Sulfur oxides (SO <sub>x</sub> )**	0.03 ppm	Annual arithmetic mean*	3 h: 0.5 ppm
	0.14 ppm	24 h	= 1300 µg/m <sup>3</sup>

\* Pollutant concentrations are distributed geometrically. Using the arithmetic mean rather than the geometric mean biases the estimate of central tendency upward, so that the standard is more, rather than less, protective.

\*\*Sulfur oxides are the only pollutants that retain a secondary standard different from the primary standard. The secondary standard is 0.5 ppm averaged over 3 hours.

These standards have been revised many times and are now under sequential review following a timetable established by EPA.

Each NAAQS has two levels: a primary ambient air quality standard and a secondary standard. Primary standards are set for the protection of human health. Secondary standards are set for economic protection (for example, to prevent crop or tree damage, or in response to a particular problem with ozone and sulfur oxides), visibility (primarily an issue with coarse particulate matter), and other aspects of the “public welfare.” In reality, the concept of secondary standards has become increasingly meaningless because the standards have been dropped or set equivalent to the primary NAAQS for all but one pollutant: sulfur dioxide. The timing of the biological action of sulfur dioxide required a standard that restricted peak exposure in terms of a maximum allowable over 3 hours; the primary standards for 24 hours and for an entire year averaged out peak levels, and therefore were not protective.

Lead is a legacy of leaded gasoline and so may still be an air toxic in developing countries. Air concentrations of lead have declined since the 1980s. In 2008 the NAAQS was reduced to recognize the neurocognitive effects of lead at low exposure levels, and the absence of a threshold.

The EPA also regulates air toxics, on a pollutant-by-pollutant basis.

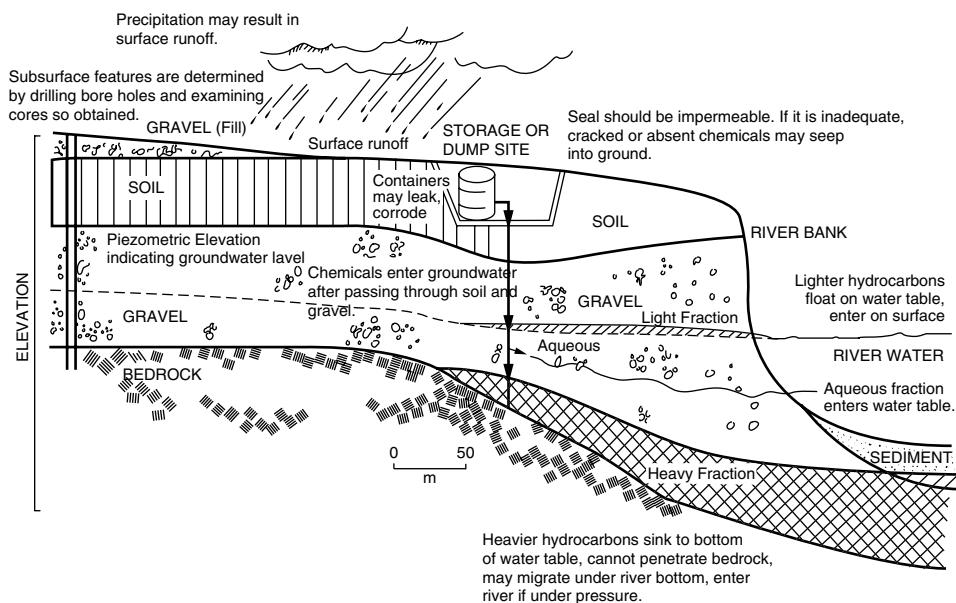
## **WATER QUALITY**

Water resources must be suitable in quality, quantity, and access to protect human health. All water resources are part of the hydrologic cycle, in which water from the oceans and inland bodies evaporates, and then is precipitated on land as rain or snow to form streams, lakes, rivers, and percolates into the ground to form “groundwater.”

“Surface waters” are easily contaminated, but if the flow is adequate they can eventually dilute and displace chemical and microbial contamination. All water flows downhill by the force of gravity unless there is a pressure gradient of equal force in another direction. (The beauty of hydrology is that one simple principle explains so

much.) These bodies of water eventually flow to the ocean or, in arid regions, to closed bodies of water that lose volume by evaporation. An extreme example of the latter is the Salton Sea, in southern California, where drainage from agricultural areas and salty soils in the desert flow into the lake. Concentration by evaporation there has resulted in very high salinity and high concentrations of pesticides and other chemicals.

“Groundwater” is much more difficult to manage. Figure 12.2 illustrates a cross section of land near a river. Rain and other precipitation passes through the soil and the strata underneath the topsoil, a process called “percolation.” It passes through clay slowly, and through sand or gravel very quickly, eventually reaching the underlying rock strata. Often there will be deposits of sand, called “sand lenses,” through which water can move more quickly than through surrounding clay or soil. If there are fractures in the rock strata, water can pass through them more quickly. Otherwise, the water collects in



**Figure 12.2.** Groundwater movement.

pores in rock or between grains of soil or sand. Eventually in its descent the water reaches a relatively impermeable stratum of rock called the “aquitard,” which stops its descent. The water may move in a horizontal direction by the pull of gravity and the path of least resistance; if it cannot, it will pool in the porous rock above the aquitard. Such large pools are called “aquifers” and are tapped by wells. If the underlying water is under pressure because of the weight of the water above it, the water will flow upward in a well; this is an “artesian well” or a spring. Otherwise, the water may discharge into a waterway (as in the figure) or collect underground, recharged by rain at the surface.

Contaminants on the surface are carried downward by water percolating through layers of soil and rock. Contaminants that are lighter than water float on top of the groundwater, and those that are heavier sink, just as they do in surface water. Many contaminants also bind to soil particles, especially to clay, and some are bio-transformed by bacterial or mold metabolism, or by chemical changes as they percolate. The material that reaches groundwater is not identical to the contamination at the surface. Groundwater is usually cleaner than surface water because it has been filtered during percolation through overlying strata of soil; but once contaminated, groundwater is much more difficult to decontaminate than surface water. Contamination of large aquifers is essentially permanent. In vast areas of North America, aquifers are the only practical available water source. These aquifers have been extensively tapped, and major aquifers in arid regions are now seriously depleted, contributing further to the problem of access to clean water.

### Drinking Water Disinfection

Water resources management involves assurance of a secure and reliable supply of water and treatment of drinking water to guarantee safety from microbiological and chemical hazards and to ensure acceptability by the public. The technologies of water distribution and treatment and of wastewater treatment and discharge are not

complex, but they must be extremely reliable, and the cost of maintenance is very high. Unfortunately, many urban water and sewer systems in the United States have been neglected due to financial constraints affecting local government.

Purification of drinking water can be accomplished in many different ways, but filtration and chlorination are the two methods used on a large scale to produce “finished water” for distribution. These techniques are consistent, reliable, and cost-effective. Together they have dramatically reduced the incidence of water-borne diseases to their current low levels. Chlorine, in its various forms, has the great advantage of continuing to work downstream, so that minor breaks in the distribution system, with low levels of contamination, are effectively disinfected all the way to the tap. Other methods of disinfecting water, such as ozonation, do not do this, and although they may be effectively bactericidal in the first instance, they do not provide downstream protection. This is the major reason that chlorination continues to be the standard for water disinfection. In order to reduce the amount of residual chlorine in the system while maintaining its protective effect downstream, most utilities are now switching to chloramines. These are a mix of similar compounds that break down into ammonia, for an instant but not persistent bactericidal effect of initial disinfection, and chlorine, which has a persistent effect downstream. Chlorine dioxide, which works by slightly different principles, is also effective but much more expensive.

## **Wastewater Management**

Wastewater management involves the handling and treatment of wastewater so that it can be safely discharged and avoid contact with drinking water or food.

Wastewater treatment is categorized as primary, secondary, and tertiary. In primary treatment, the sewage is held in settling tanks and the particles are allowed to sediment. Organic material is digested by bacterial action, and the sludge that forms is eventually dried and disposed of separately. Secondary treatment allows further bacterial

degradation of wastewater organics in an oxygen-rich environment. Tertiary treatment involves chemical purification of the remaining bacteria and organics before the effluent is released, usually into a river or ocean. Most sewage treatment in the United States is primary. Occasionally, raw sewage is still released without treatment, especially when primary treatment facilities are out of order or a flood has overloaded the combined sewer lines, in cases where the line holds both sewerage and flood water. Flood channels and sewer lines are no longer built to interconnect in modern cities, as they once were for purposes of flood control; the early sewers of London were adapted for sewage treatment after they were originally built to control flooding. However, many older cities in the United States still have “combined sewers.”

Industrial wastes are putting a new burden on wastewater treatment facilities, because chemical contamination is not well handled by conventional treatment designed to remove sewage pathogens and organic materials. On occasion, a surge of chemical contamination in a sewer line may be bactericidal, killing the bacteria that are essential to primary and secondary treatment, and thereby destroying the effectiveness of the sewage treatment facility. There is great concern today that the commercial introduction of products containing antibacterial silver nanoparticles will degrade wastewater treatment plants because these particles pass through treatment easily and are bactericidal.

### **Microbial Hazards**

Within limits, a flowing body of water is a self-purifying medium insofar as bacteria and viruses are concerned; it can dilute, oxidize, and remove pathogens as long as its capacity is not exceeded and sufficient time elapses before water is withdrawn downstream. Where population pressures and the density of urbanization place heavy demands on water resources, however, the self-purifying capability of water is rapidly outstripped, and the same body of water may become a vehicle for the distribution of pathogens. In many older cities or

among cities along a river such as the Mississippi, the intake for drinking water may be just downstream from a sewer outfall, placing a tremendous burden on the efficiency of water treatment facilities to protect human health. Fortunately, the disinfection process is very effective in this respect.

The type and degree of pathogen entry into water depends on the nature of the contamination and the health of the population from which the sewage in wastewater originates. In the United States, outbreaks of water-borne disease are usually limited to enteritis associated with any of a number of viruses (including norovirus), as well as the parasitic disorders cryptosporidiosis and giardiasis. Hepatitis A, B, or C are uncommon, with hepatitis A most frequently encountered. Typhoid has been eliminated in the United States.

By far the most important water-borne illness in the United States is diarrhea. Young children are especially vulnerable to dehydration and are most at risk for serious illness. Most cases of water-related diarrheal disease go unreported and may be easily confused with food intoxications or transient influenza. Even when such outbreaks occur, they rarely cause life-threatening illness. In less than 50 percent of the cases, the etiologic agent can be identified in retrospect. Prospective and case studies of diarrheal disease strongly suggest that many different viruses are involved.

The major test for bacterial contamination in water is the “coliform count,” a robust test that demonstrates the presence of coliform bacteria, but not necessarily human *E. coli*. It is a reliable indicator of fecal contamination, and a first alarm for the presence of water contamination. Further tests can be done for fecal *E. coli*, enterococcus, and other specific pathogens. The familiar agents responsible for food contamination (including *Salmonella*, *Shigella*, and *Campylobacter* spp.) can also contaminate water.

Surface water is more easily contaminated than groundwater. However, there are conditions in which microbial contamination of groundwater can be extreme. In 2000, in the small town of Walkerton, Ontario, runoff from a feedlot contaminated a well that

was incompetently managed and insufficiently disinfected, leading to an outbreak of *E. coli* O157:H7 that killed 7 people and made 2,500 people ill.

### Recreational Waters

Disease outbreaks that are due to exposure to pathogens or chemicals in “recreational waters” (usually via human diarrhea) are less common than incidents associated with drinking water, simply because of the circumstances of exposure and the potential for illness. Recreational water (a misnomer, because it refers to human contact with any surface water other than drinking water, including ocean water) is regulated by different EPA rules that emphasize microbial contamination. Coliforms are tolerated at low counts. The risk of exposure to the same pathogens discussed above may be present in recreational waters, with other possibilities, including leptospirosis (particularly where rats are present).

What one might call “treated recreational waters,” such as swimming pools and hot tubs, may pose special hazards if not kept scrupulously clean and disinfected. Amebiasis (from *Naegleria*, a particularly nasty ameba that causes meningoencephalitis) has occurred in outbreaks attributed to hot tubs. *Pseudomonas* outbreaks can occur if the residual chlorine levels in hot tubs and swimming pools are not adequate.

Unregulated waters for irrigation and other uses may present risks in other parts of the world. Schistosomiasis, for example, is a risk in endemic areas, most recognized in the Middle East and northeastern South America, as are parasitic infestations from liver and lung flukes, most commonly in east Asia.

### Chemical Contamination

Contamination of drinking water by toxic substances is an old problem. The major chemical contaminant of surface water and wells (tapping into groundwater) historically has been nitrate, mostly in

the form of agricultural runoff from fertilizer application. Nitrate is not very toxic to adults, but is a serious threat to infants until about age six months. Infants are most susceptible for many reasons. They lack the capacity to secrete acid in the stomach at adult levels. The higher pH in the stomach and gut allows growth of bacteria that bio-transform nitrate to nitrite in the stomach. Fetal hemoglobin (Hb F), which is still a large fraction of an infant's hemoglobin, is also more easily converted to methemoglobin than is adult hemoglobin (Hb A). An infant also consumes much more water in proportion to body weight than an adult does.

Arsenic from natural sources is a serious contaminant of well water, and is associated with increased cancer risk. Although the problem is best known as a local or regional problem, in some parts of the world (such as Bangladesh, Taiwan, and Chile), sources of arsenic place thousands of people at risk elsewhere. Arsenic-containing groundwater is found around the world, and because the deposits of arsenic are spotty under the ground, nearby wells may have very different levels of arsenic.

Intrusion of potentially toxic substances into groundwater has occurred in many communities as a result of leakage in supposedly secure disposal sites. For example, groundwater in California's Silicon Valley and wells near Tucson, Arizona, have been contaminated with trichloroethylene (TCE), an industrial degreasing agent. One particularly widespread problem has been the leakage of methyl-tertiary-butyl ether (MTBE), a fuel oxygenate suspected to be carcinogenic, from underground gasoline storage tanks. MTBE travels quickly and efficiently through water (a feature formally known as "fugacity") and has contaminated groundwater in many locations across the United States. MTBE has been banned in the United States (it was never used in Canada) and has been replaced by ethanol.

Disinfection may result in the formation of "disinfection by-products" (DBPs), which are associated with their own set of problems. In the presence of small amounts of organic material from the source water, the process of chlorination produces small quantities of potentially carcinogenic compounds called trihalomethanes.

Concern over the carcinogenic potential of the trihalomethanes must be tempered by the observation that the demonstrated carcinogenicity of some of the compounds requires exposure levels several orders of magnitude higher than their concentrations in water. There has also been concern over the possible role of DBPs in reproductive toxicity, primarily in the induction of miscarriages. This has now been studied extensively, and the evidence suggests that it is not a risk.

Lead in drinking water is an old, historically well-documented problem. Lead is a problem of sources in the home and of a lead service line, if present; it is not usually found in source water. The water delivered by water mains should be essentially free of lead. Lead in drinking water is caused mostly by the solubilization and release of lead in home fixtures (such as faucet valves, solder on pipes, and lead-containing water meters) and lead water lines, which were used extensively until the 1970s to bring water from main lines into the house. (Larger buildings used stronger materials, usually cast iron.) Modern and renovated houses have lines made of other materials, mostly copper and PVC. Lead release occurs because of water chemistry changes that affect the inner surface of the pipe, and is a particular problem where the water is very soft. In 2004, due to concern over the formation of trihalomethane reaction products in water (see above), a rule called the Disinfection Byproduct Rule mandated a change in management of the disinfection of drinking water. The objective was to reduce the residual chlorine levels in water in order to reduce reactions with organic material from the source water that produced DBPs. Large utilities mostly switched to chloramines, an effective formulation that relies on ammonia to reduce the residual free chlorine level but maintains sufficient chlorine for downstream protection. Some cities experienced changes in water chemistry leading to increased solubilization of lead from fixtures in homes and from lead service lines still present in older homes. The increase in lead exceeded EPA regulations and caused local problems, attracting a great deal of attention, especially in Washington, D.C. This problem was solved by the addition of phosphoric acid, a familiar and widely used water and food additive, which “passivated” (stabilized) the

surface chemistry of lead-containing materials and prevented further release. A program to replace lead service lines was also introduced. The issue became well known as an example of the problem of “simultaneous compliance,” in which utilities or other regulated bodies sometimes have to deal with the unintended consequences of incompatible regulations. There is no maximum contaminant level (MCL) for lead in the United States because it is not a problem in source water. Canadian provinces do have a drinking water standard for lead as does the World Health Organization (WHO) (both 10 ppm), with a guideline for continuously flowing water, which rarely has high lead levels. U.S. EPA regulations are based on an unusual standard called the “Lead and Copper Rule” (LCR), which seeks to control the corrosivity and release of lead from the distribution system as a whole. The LCR is based on what might be called a “near worst case” scenario that represents an extreme of the distribution system rather than a representative sample of water. The 90th percentile of homes tested, following a detailed protocol that all but maximizes lead levels, must fall below 15 ppm. This has been misinterpreted as an MCL (for example, one CDC Web site refers to this level as if it were a health standard), but it is not.

Since 2003, a series of reports from the U.S. Geological Survey reported detectable but low concentrations of pharmaceutical agents in surface waters: lakes, rivers, and even tap water in major U.S. cities. The pattern of detection closely mirrored the frequency of use of these pharmaceuticals: contraceptive estrogens, non-steroidal anti-inflammatory drugs, anti-seizure medications, antibiotics, and so on. One source of these chemicals is upstream wastewater, excreted from patients, and on occasion from disposal of excess medication. Other sources are out-of-date medications dumped into the toilet or down the drain, which are only partly removed by conventional wastewater treatment, and the agricultural use of medication, resulting in animal antibiotic excretion. Although the levels are much lower than would be expected to affect human health, the finding received a great deal of public attention and generated much alarm. There is convincing evidence that an ecotoxicological effect does exist, however. Chemical

contamination with pharmaceuticals, as well as other organic chemicals and metals with endocrine-disrupting effects, is associated with abnormal sexual development of fish, whose aquatic habitat makes them highly susceptible. There is also concern that chemical contamination levels will increase with a growing and aging population.

Chemical contamination of drinking water is an issue that is growing in importance. Prevention and careful handling of hazardous material incidents is the key to dealing with this problem. Decontamination of water resources, particularly groundwater, is difficult, expensive, and, for low concentrations of chemicals, often impossible. Most water contamination incidents involve a small number of industrially important chemicals, such as solvents and pesticides, and metals. In recent years, a new set of chemical pollutants has been identified and given the name “emerging contaminants.” These chemicals have received much attention but are not yet regulated by the EPA. The most prominent include the following:

- *N*-nitrosodimethylamine (NDMA), a nitrosamine (and, like most of the class, a suspected carcinogen) that is a rocket fuel, an air pollutant, and a natural product found in food. NDMA is highly miscible in water and behaves much like MTBE.
- Perfluorooctanoic acid (PFOA or C8), a suspected carcinogen, is a synthetic constituent of Teflon and related fluoropolymer products and is found near industrial sources.
- Perchlorate is a suspected carcinogen that is present as a natural brine salt but is also produced in the manufacture of rocket fuels. In a controversial decision of 2008, the EPA declined to regulate it despite recommendations that it do so, and it has been alleged that this decision was in deference to the Department of Defense. This is an example of how politically fraught the effort to deal with emerging contaminants can be.
- 1,2-Dioxane (not to be confused with dioxin) is an IARC-identified, probable human carcinogen, apparently acting by epigenetic means, that has been in common use as a solvent.

- 1,1,1-Trichloropropane (TCP) is a recognized (by California) human carcinogen used as a solvent and chemical feedstock. It contaminates groundwater.
- Polybrominated diphenyl ethers (PBDEs) and polybrominated biphenyls (PBBs) are IARC-identified possible carcinogens and may carry a risk of thyroid toxicity. They were formerly used as fire retardants but are no longer produced. They remain as persistent organic pollutants, with sediment as a depot in bodies of water. They are homologous to their more familiar chlorinated counterparts, but like most brominated homologues they are more toxic.
- Silver nanoparticles are new bactericidal components primarily found in textile products. They pass through filters easily and have the effect of killing bacteria in wastewater treatment plants, rendering the plants less effective.

## SOLID AND MUNICIPAL WASTE DISPOSAL

In general, physicians have had limited interest in solid waste disposal in the past. Other than vector control, landfills, especially, have not been major threats to human health. OEM physicians who have responsibility for environmental health management, however, should know that landfills and related site issues are among the most common problems they are likely to face.

The amount of garbage and trash produced by humans is increasing. Commercial and domestic solid waste is a great practical problem for many local governments. Because of population distributions, land use restrictions, transportation costs, and concern from society over environmental effects, there is intense pressure to find alternative solutions. This has led to increased interest in methods such as source reduction, recycling, waste-to-energy plants, and landfilling.

The “not in my back yard” (NIMBY) syndrome makes it very difficult for local governments to find new sites for subjectively undesirable developments such as landfills. Citizens understand that a

solution to the solid waste problem is necessary but few are willing to accommodate the solution.

Solid waste is generally made up of five components, listed in Table 12.6. The terms used to describe these various components are universally agreed upon by solid waste technologists. The garbage component tends to provide the nutrients and breeding grounds for flies and rats, whereas the rubbish and debris provide the housing. Buried solid waste can create ground and surface water pollution. This problem, however, is minuscule compared to the problems created by untreated and undertreated municipal sewage, and by effluent of industrial and agricultural wastes that reach water sources.

Most municipal solid waste systems divert chemical and other hazardous wastes so that they do not contaminate the solid waste stream. These hazardous wastes are required to be disposed of in a secure receiving site complying with EPA regulations. Even so, small amounts of domestic and commercial hazardous waste, such as discarded cans of insecticide or paint, may find their way into the trash stream.

**Table 12.6.** Typical Composition of Domestic and Commercial Solid Waste

Component	Characteristics	Origins
Garbage	Putrid; attracts rodents and flies; suitable for hogs and animal feed recovery	Food wastes at all levels of handling
Rubbish and debris	Non-putrid; organic and inorganic; combustible and noncombustible	Packaging; grass, shrub, and tree clippings; leaves
Ashes	Inorganic inert from combustion	Coal and wood burning; solid waste incineration
Street sweepings	Mixed; high in paper, grit, and dust	Litter, abrasion, spillage, dogs
Oversized discards	Combustible and noncombustible; some parts can be salvaged	Used furniture, home appliances, motor vehicles

Generally speaking, there are four methods of disposal on a large scale: incineration, open dumps, sanitary landfills, and composting. All begin with the assumption that hazardous wastes are diverted out of the waste stream. Hazardous waste is discussed in the next section.

An incinerator, fueled by wastes, reaches temperatures of 700–815°C, high enough to burn garbage and rubbish but not high enough to dispose of hazardous organic waste. Since an incinerator can operate 24 hours a day, this method can be very efficient for a high-volume waste stream. However, although incinerators can be used for cogeneration of power, they are associated with high operational and maintenance costs. Since ash, cans, and bottles are not combustible, further handling is needed—a disadvantage of this system of disposal. There is also a problem of particulate and odor emissions in a waste holding area, which may result from a poorly operated incinerator. The major concern with incinerators is with emissions to air, including air toxics, metals, and carbon dioxide loading.

Open dumping offers no advantages other than the fact that it is an inexpensive method of disposal. There are, however, many disadvantages. This method provides optimum breeding conditions for flies and rats; mosquito larvae habitat in trash that collects standing water; and air and water pollution resulting from dump fires and leaching, respectively. Existing dumps are often tolerated by local government, but new ones are not allowed.

Sanitary landfills are the modern standard for waste disposal. In this method, a bulldozer spreads and compacts waste in chambers lined with a clay and (usually) plastic liner to prevent migration of “leachate” (percolated water passing through the trash) into groundwater. The leachate is collected by a drainage system and pumped into a wastewater treatment facility in large systems. The chamber is covered with soil and sealed. Thorough compaction of waste to prevent air pockets prevents landfill fires and odors. Aerobic decomposition occurs within the landfill, releasing methane. The methane must be captured and burned or used as fuel in larger systems, both for energy efficiency as well as to prevent a fire hazard. After closure of the landfill site, the land may be reclaimed for restricted use.

Properly managed, sanitary landfill sites present a relatively low health hazard. Workers are subject to injuries and skin disorders from contact with solid waste in landfill material, making personal protection desirable, and are subject to the same occupational hazards as general material handling and minor construction work. However, if medical waste is diverted, landfills are not high-risk workplaces. Microbial aerosols can arise from the microbial content of refuse. This phenomenon has not received much attention. Airborne total and fecal coliform counts of about  $10^3/m^3$  have been observed at some landfills. Health evaluation of workers in confined areas may be warranted, but there is no evidence of off-site health impacts from microbial aerosols as a general public health problem.

The principal concern for long-term health effects comes from the possibility of uncontrolled disposal of hazardous substances on an unapproved site. This does not seem to be a serious or unmanageable problem in practice, however.

Composting depends upon the aerobic decomposition of waste by the action of bacteria and fungi. This method of disposal can be very advantageous for several reasons. It conserves and recycles waste. It provides humus for soil and is free of pathogens. After decomposition, it is half the weight or volume of the original material, and it provides only minimal attraction to flies. However, composting on a large scale presents several problems. It may put traces of toxic metals and residues in the final product, and the composting reactions can be inhibited if ecotoxic chemicals find their way into the composting material. Composting requires presorting of wastes, grinding and turning during the decomposition process, and high equipment and maintenance costs. In addition, the compost produced requires a market and involves a transportation expense. The non-compostable wastes in the waste stream require a separate system of disposal.

Two other methods are appropriate for small-scale or selective disposal: garbage disposals, which grind organic matter and add it to sewage, unfortunately resulting in increased chemical oxygen demand; and salvage or recycling, which mostly involves conversion and re-use of metal and paper products recovered from the track stream.

Volatile emissions arise with any waste source containing volatile organic compounds. Normal domestic refuse does not have substantial volatile content, although odorous volatiles can be created by decay of putrescibles (garbage that putrefies rather than decays, such as raw meat). Possible sources of volatiles are various organic solvents that may be stored at the home or discarded from a business; their disposal to a municipal landfill would be illegal in most jurisdictions. Odorous emissions from putrefaction are generally controlled by frequent covering of refuse with clean soil.

Fires at landfill and dump sites have caused serious air pollution and odor problems. They can be very difficult to put out if they become established in the fill itself. Continuous vigilance is necessary to screen waste deliveries to prevent the dumping of heated or chemically reactive materials.

## **“CATEGORICAL” POLLUTION AND HAZARDOUS CHEMICALS**

Categorical pollutants are chemicals that are managed or treated individually, one at a time, as opposed to “media pollution” (air, water, soil, food, and consumer products), in which an integrated approach is normally used. Air toxics, discussed in the context of air quality, are categorical pollutants.

### **Pollution**

One definition of pollution is the production of a substance in quantities in excess of the natural cycle’s ability to assimilate or convert the excess, such that accumulation occurs and a biological response results. In the case of carbon dioxide in air or nitrates in water, huge quantities of the substance must be accumulated for an effect to be detectable, and that effect is not biological in the first instance (its consequences certainly are, however). In the case of lead or a highly toxic compound such as dioxin, it may only take a small amount, depending on the capacity of the ecosystem to absorb or

dilute the substance. This definition implies that every pollutant can be characterized on the basis of production, degradation, distribution, and relative toxicity, much as one characterizes the pharmacology of a new compound. In practice, much is unknown even about familiar chemicals, and there is far greater uncertainty about each of the 800–1,000 new chemicals introduced every year.

The Toxic Substances Control Act (TSCA), which was the primary vehicle for chemical regulation in the United States, formerly the world's largest producer of chemicals, has been largely supplanted as the international standard by the European Union's REACH program (Registration, Evaluation, Authorization and Restriction of Chemicals), which has created the European Chemicals Agency. Producers and distributors who now seek to sell to a world market must comply with the more stringent REACH requirements because TSCA is no longer sufficient.

In the 1980s there was a strong movement to democratize the response to chemical hazards by making information more freely available to citizens and communities. The result was a series of federal laws and programs that put detailed emissions and pollution information within reach of the public, and ultimately on the Internet. After 2001, much of this information has been taken off line because of its value to potential terrorists.

The EPA has developed the Environmental Justice Geographic Assessment Tool (<http://www.epa.gov/enviro/ej/>) as its primary means for dissemination of local information. This Web site allows the user to specify the state, city, watershed, zip code, or GIS position of interest to bring up a map of the area with locations of regulated sites (such as regulated industries or Superfund sites), healthcare facilities, transportation infrastructures, waterways, air quality monitors, water quality monitoring stations, schools, and churches. The maps can be overlain with demographic data, for the assessment of environmental justice.

The National Environmental Monitoring Inventory, which is still a work in progress, is an initiative to achieve a comprehensive ecological assessment of the United States, with a site that ultimately will

link databases on media quality and emissions from federal and state sources. It would include monitoring data from many sources supplemented by a variety of surveys, studies, and remote sensing data sets as they become available. This is a formidable challenge in data management, but an even greater challenge for assessment, even simply correlating data from these disparate sources.

The Toxic Release Inventory (TRI) is an effort to capture emissions from the private sector and other major sources into a database at the community level. It was mandated by the Emergency Planning and Community Right-to-Know Act (EPCRA) in 1986, following the catastrophic incident in Bhopal, Madhya Pradesh (India), in which a Union Carbide plant released highly toxic methyl isocyanate during upset conditions, resulting in many thousands of deaths. The TRI is heavily used for risk management, for advocacy, and to support litigation. Data from the TRI indicate that the U.S. chemical industry has reduced emissions by more than 70 percent since 1987.

The Resource Conservation and Recovery Act of 1976 (RCRA), one of the older foundational pieces of legislation for hazardous materials, requires “cradle to grave” tracking of hazardous products through production, transportation, and end use. Whether through necessity or conviction, the chemical industry has embraced this mandate and adopted the concept of a life cycle approach to product stewardship up to the point where it loses control over the end use.

Many sites with environmental data, together with sites presenting health and health risk data, can be accessed from one Web page maintained by the Centers for Disease Control and Prevention (CDC): <http://www.cdc.gov/nceh/data.htm>. Otherwise, CDC’s role consists of the public health protection activities of the Center for Environmental Health and the legislatively separate but functionally integrated Agency for Toxic Substances and Disease Registry. The latter has the mandate to prepare documentation on individual pollutants and toxic chemicals (in the form of the essential series *Toxicological Profiles*), and to investigate situations of community exposure, such as among residents near Superfund sites.

The EPA has also placed great emphasis on initiating community advisory committees and encouraging stakeholders, such as large companies, to do the same. This is a legitimate way to get community input on pollution risks, but it is also a way to bring activists and community representatives (not always the same people) into a relationship with the agency that marginalizes more extreme groups.

Canada maintains similar individual databases, not yet linked.

### Hazardous Chemicals

The presence in the 1970s of numerous toxic waste disposal sites in the United States galvanized public opinion and led to a legislative response that created Superfund, as described below. It also initiated a thirty-year search, largely unproductive, for epidemiological evidence for health effects related to proximity to hazardous waste disposal sites. In order for chemicals to have an effect on human health, there must be an “intact pathway of exposure” by which the chemical is conveyed to the person, whether by air, surface water, groundwater, or soil contact. With exceptions, an intact pathway of exposure is usually absent, which accounts for the many negative or unconvincing studies. Even so, these sites represent the potential for release, for human exposure under unusual circumstances, for occupational exposure, and for ecotoxicological effects, the latter described elsewhere in this chapter. Thus, the problem of disposal of toxic substances is not overrated but merely underestimated in its complexity. Human health risk has simply not been the predominant threat.

Toxic substances are not new. Exposure to potentially dangerous chemicals has occurred in the smelting of metals and the tanning of hides since ancient times; hydraulic mining released heavy metal into surface waters and groundwater during the nineteenth century. In the early days of industrialization, urban areas were saturated with pollution, including chemical wastes. (Other than metals, however, most of this contamination was biodegradable.) The contemporary problem of hazardous waste reached a greater degree of urgency in the late twentieth century because of several historical factors.

The number and hazardous nature of toxic substances in common use changed dramatically after the Second World War. Research and development in organic chemistry and chemical engineering resulted in thousands of new compounds entering widespread commercial use, including such persistent compounds as the polychlorinated biphenyls (PCBs), more potent pesticides, accelerators, and plasticizers with unusual and poorly understood effects. Current control efforts are dealing with the backlog while faced with a continuing flood of new products. Nanomaterials are about to change this again, by reformulating and redesigning formerly familiar materials in new ways in which their chemical properties, physical behavior, and biological activity are much different than previously seen in the bulk material. The scale and design of nanomaterials confer new health risks, in most cases not easily predicted by the conventional toxicology of their constituents. Consequently, much that is familiar and settled in toxicology and chemical hazard evaluation will have to be revisited.

A stream of new chemicals enters commerce at the rate of 800 to 1,000 per year, according to data cited by the National Academy of Sciences (2007). The production of chemicals has also risen dramatically, and manufacturing has spread across the globe. In 1941, production of all synthetic organic compounds in the United States alone was less than 500 million kg, and the United States was the dominant producer. Today it is estimated to be much greater than 100 billion kg, and the United States is no longer dominant in production. Robust chemical industries have developed in many developing countries, such as China and India, with the result that there is much more product in circulation. Although the technology for producing chemicals has changed to emphasize more efficient, contained, and emission-free production, plants in developing countries often choose obsolete but cheaper technology and equipment. Thus, in terms of sheer volume, the nature of chemical products, the technology of production, and the distribution of production, the chemical industry is not necessarily a different world than in the past. Rather, it is a case of the future coexisting with the past, with increasing diversity and challenges of keeping up.

Toxic chemicals are much more intrusive into daily life. Many chemical plants or disposal sites that were once isolated or on the edge of town have become incorporated into urban areas by suburban growth. Communities now lie in closer proximity to the problem than they have in the past. Some communities were built directly over old disposal sites, with economic and political consequences in the case of Love Canal, a seminal case in point from the 1970s.

### **Chemical Safety**

Chemical safety is a general term for managing the hazards and risk associated with chemicals in order to protect workers and community residents from all levels of hazard, but in practice it usually refers to managing emergency conditions from incidents of uncontrolled release or exposure. Most chemical safety issues involve uncontrolled release during upset or emergency conditions and spills during transport. The primary concerns associated with particular chemicals may include the risk of explosion, flammability, acute toxicity, release of heat (exothermy), chemical reactivity, and corrosiveness. (Asphyxiation is a risk in confined spaces.)

The Emergency Planning and Community Right-to-Know Act of 1986 (EPCRA) introduced several measures to strengthen emergency planning and community response and cooperation in the event of chemical incidents. It required community response plans for facilities and transportation corridors with a defined risk of chemical exposure, mandated involvement of emergency planning groups, required access to hazard information (for example, community access to MSDSs), and encouraged community participation through consultation and stakeholder participation. However, many of its provisions were cut back, and implementation was rescinded after 2001 amid growing fears that detailed knowledge of facilities and planning could play into the agenda of terrorists.

Chemical safety in practice involves protecting residents who live and work in communities where chemicals are made, stored, or used or through which chemicals are transported. It implies involving

public officials in planning for emergency response, and communicating with the community, involving them early in planning for incident response. Several elements are common to such plans:

- Stakeholder representation and involvement in community planning.
- Data gathered on who lives near potentially critical facilities (highways, railroad tracks, pipelines, gas wells) and may be affected.
- Identification of susceptible populations. Except in sparsely populated areas, where people can be approached one on one, it is usually impractical to ascertain the circumstances of individual residents. However, the locations of schools, hospitals, day-care centers, assisted living facilities, and other concentrations of residents at exceptional risk are important and helpful in orientation to the community.
- Communication strategies. Whether by telephone tree or mass cell phone messaging (which has been introduced in Finland and on a sign-up basis in Washington, D.C.), a system of communicating with residents in a crisis is essential. Ideally, this should be coordinated with public health officials and should cover all threats, but large companies often have more resources and motivation to put a system in place.
- Training in “shelter in place.” The basic strategy for uncontrolled release of airborne chemical hazards in the community is called “shelter in place” and involves residents staying indoors with windows closed and HVAC or other ventilation systems turned off. The residents are instructed to stay in place until they receive word that the airborne hazard has dissipated or has passed over the area and moved far downwind. Because concentrations of potentially toxic gases rise only slowly within the contained space of an unventilated building (while the contaminant plume itself is moving outside, depending on the wind), this strategy is very effective. Impermeable adhesive tape (duct

tape) and plastic sheeting may be useful for sealing windows, doors, and other openings under stagnant conditions but is unlikely to be necessary, as the incident is likely to be over by the time such action can be completed.

- Household emergency planning. Kits for emergencies should be part of the equipment in every home; therefore, this is a general measure for emergency management. Kits should contain the essentials for living through the first two or three days of a disaster (sufficient water, food, candles, flashlight, a radio, replacement batteries, medicines, duct tape, etc.), but the nature of chemical incidents is such that they would not last this long.
- Evacuation planning. Evacuation is generally a poor response to a chemical incident once it is in progress, since changes in wind may place evacuees who venture outdoors at risk, and congestion on highways can lead to a dangerous situation. However, if a risky situation is developing slowly and there is sufficient time, evacuation in a controlled manner may make sense.

Most uncontrolled chemical releases occur during transportation incidents, such as derailed railroad tanker cars or trucking accidents. Many of these incidents involve multiple failures or a sequence of events. For example, on November 10, 1979, 250,000 residents of the city of Mississauga, Ontario, had to be evacuated when a train derailed. The derailing triggered a series of “boiling liquid expanding vapor explosions” (commonly called BLEVEs in the field of transportation safety) in three tanks carrying propane, which resulted in the puncture of a tank car containing chlorine. Fortunately, nobody was killed. The worst chemical disaster from a fixed site was the infamous incident in Bhopal.

The chemical industry instituted a program called Responsible Care in 1985, requiring participating companies to practice stewardship of chemicals throughout the product life cycle, community participation, planning for emergencies, emissions reduction, occupational safety, and environmental sustainability. Now worldwide,

Responsible Care has also addressed community concerns about living next to chemical plants.

Chemical safety since 2001 has emphasized prevention of opportunistic acts of terrorism on site or the commandeering of hazardous materials for intentional assaults. This represents a shift in point of view from the chronic, long-term exposure paradigm that had recently dominated views of chemical safety back to a more traditional focus on hazard: flammability, explosive risk, chemical reactivity, and short-term acute toxicity.

It is well recognized that substituting less hazardous chemical substances, shortening the supply chain to reduce transportation incidents, and containing production processes reduce the risks associated with chemical facilities.

RCRA required that spills of defined “reportable quantities” (RQs) of toxic chemicals be reported immediately to a National Response Center (NRC) managed by the EPA, which now receives reports of spills under other legislative mandates, including EPCRA, the Clean Water Act (section 311), the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), and the Hazardous Materials Transportation Act of 1974. NRC serves as a single window for reporting purposes, discharging responsibilities under any and all of these acts. However, state agencies must also be notified as appropriate, unless the spill is confined to the responsible party’s own site.

The U.S. Chemical Safety and Hazard Investigation Board (CSB) is an independent commission authorized by the Clean Air Act Amendments of 1990. It performs highly detailed investigations of chemical incidents, including spills and on-site fatalities, for purposes of establishing the root causes and identifying contributing factors. In this it is quite different from other investigative bodies, which are normally tasked with assigning responsibility and establishing liability, or in support of enforcement. CSB probes deeply, without regard for rulemaking, litigation, or determining regulatory compliance, and thus can focus on reconstructing events and providing lessons for prevention. It is a rare example of a government agency that is

designed to learn from history and to make recommendations to prevent negative events from occurring in the future.

### **Household Exposure and Community Collection Programs**

Exposure in the home is perhaps the leading means by which children come into contact with toxic substances. Exposure to a toxic substance may occur at several points in the use cycle of the substance. People work in a plant that uses the substances in an industrial process, and they may fail to change clothes or wash before coming home. They may reside near hazardous waste disposal sites that are illegal or poorly designed, or that allow opportunities for exposure from accidents or careless handling. Additionally, these sites may be deficient in the containment of the substance or in the use of fencing to keep children off the site. Exposure may also occur in the home as the result of mislabeled, poorly stored, and non-childproof consumer products. Consumer education to promote awareness of the potential toxicity of common products is urgently needed. Pesticides in aerosol cans, bleaches, household cleaners, and cleaning fluids are potentially dangerous to children and must be treated as such.

A major unresolved issue in municipal solid waste handling is contamination by hazardous waste disposed of by accident or by intent. This can be minimized by diverting disposal to a separate waste stream. In the United States, people are often willing to drive up to 5 miles to safely dispose of household toxic wastes. Other means can also be used to control release of the remaining hazardous substances. Some communities have organized periodic pickup drives to collect household hazardous wastes in order to prevent contamination of sanitary landfills.

“Toxics round-up” programs and drop-off facilities to collect and properly dispose of small quantities of hazardous waste play an important role in diverting potentially hazardous substances out of the general waste stream. The individual in possession of a bottle or can of solvents, pesticides, or some unknown powder or fluid usually

cannot afford the high cost of proper disposal and does not understand the risk. A decentralized system for collecting such hazardous waste from consumers at least has a chance of intercepting it before it is poured on the ground, flushed down the toilet, or burned and released into the air. Such systems have been introduced in most urban areas, involving home or convenient pick-up of small quantities of toxic substances to be discarded. These systems are in need of subsidies to continue, however, because they do not recover their costs through recycling.

## HAZARDOUS WASTE DISPOSAL

Superfund sites are examples of how not to manage hazardous waste. Because of this experience, there is little public enthusiasm for developing secure hazardous waste disposal facilities, and much opposition of the NIMBY variety (see Chapter 8). However, well-designed hazardous waste disposal facilities, using the best available technologies of recycling, dehalogenation, and containment, are urgently needed. New disposal sites are not a perfect solution because inevitably someone has to live near one. Without new sites, however, society may lose control of the situation entirely. What happens when a hazardous waste disposal site is not available is worse than the presence of a poorly managed disposal site, as shown by the following scenarios reflecting true events:

- Waste is hauled a tremendous distance at great expense, exposing drivers and residents along the way to the risk of accidents and leakage.
- Waste may mysteriously “disappear” by the side of the road, as it has in New York, California, and North Carolina, causing exposure of residents and uncontrolled contamination.
- The waste may be mixed illegally in a landfill designed for non-hazardous solid waste, causing a serious risk of migration, local contamination, and pollution of groundwater.

- The waste may be hauled away by suspicious characters who work for cash and insist that no questions be asked. Organized crime has penetrated the business of hazardous waste handling and is a menacing threat.

The details of hazardous waste disposal are beyond the scope of an OEM physician and are within the domain of engineers. It is useful to know the principles in order to communicate with technical staff, however. The principles include chemical or biological stabilization at the source, destruction on site whenever possible, and non-transportation of liquid or gaseous material.

Thermal degradation is highly effective when feasible, but requires expensive equipment. Pyrolysis using high-temperature furnaces (usually a rotary drum design) breaks up stable organic compounds such as PCBs, leaving infinitesimal residue. Incinerators, operating by combustion, work for materials with a lower ignition or degradation temperature, but may not be sufficient for PCBs and dioxins, if they are in the waste stream. Incineration and pyrolysis have the considerable advantage of dramatically reducing the volume of wastes and sterilizing any waste contaminated with microbes. Cement plants are highly controversial but they are ubiquitous in every industrial society, achieve high temperatures, and produce few emissions. Any metal content may contribute to the product, since metals are added to commercial cement. In fact, Portland cement can bind and therefore stabilize many bulk contaminants, such as fly ash.

Physical separation (and solidification whenever possible) is critical in the management of liquid wastes, to avoid leakage of liquid or gas. Filtration through sand or fabrics can remove relatively coarse particles. Evaporation in vacuum filters, drying beds, or pans works when material is suspended in water and there are no volatile hazards. Sedimentation (in a tank or, for low-toxicity materials, in a settling pond), clarification (using agents that promote flocculation and coagulation), and centrifugation can be used to separate the solid from the liquid. This makes the solid waste easier to handle, and the liquid supernatant, if aqueous, can be put through

wastewater treatment if it does not contain highly toxic liquids. Low-density solids can be flocculated and floated on bubbles, so that they can be skimmed off the supernatant. Reverse osmosis can concentrate the solute on one side of the membrane, removing water and accumulating salts on the other, in a smaller volume.

Some gases, such as ammonia, can be absorbed by water or by solvent. Steam stripping, in which steam is injected into liquid waste, may volatilize some components and remove them from liquid solvents, leaving behind less volatile residue which concentrates dissolved solids; however, the evaporated gas must be trapped and condensed into a distillate, and treated appropriately.

Hazardous material with appropriate surface activity can be adsorbed onto media such as activated charcoal, or (particularly for radionuclides) they can be run through ion-exchange columns.

Tanks, drums, and holding vessels must be situated in secure areas, on impermeable surfaces, surrounded by a berm to capture any free liquid or contaminated precipitation, and equipped with a drain system into a lined sump or holding tank to prevent discharge into waterways in the event of a spill or leak.

Chemical treatment is limited only by the reactions possible and their safety. Neutralization of strong acids and bases is always required. Reactions that precipitate the toxic material into a solid allow use of the phase separation methods described above. Reduction is particularly useful when hexavalent chromium, a known carcinogen, is present (as it often is in electroplating solutions), because trivalent chromium is much less toxic and less soluble.

Biological methods of remediation have slowly grown more popular as systems are invented that match the capacity of chemical remediation. Bio-remediation is also particularly attractive for remediation in situ, as in Superfund sites. Bacterial degradation, under controlled conditions, may be used for bulk bio-transformation of toxic materials if appropriate strains are available; it is the principle behind composting.

When liquid wastes of relatively low toxicity are generated and it is not practical to dispose of them in other ways, “injection wells” may be used. These are deep wells, sometimes abandoned oil wells,

below groundwater levels. They have often been used in the oil industry to get rid of exhausted drilling mud (which is often high in metals), brine from wells, and process chemicals.

The best way to handle hazardous waste is to not produce it. Waste minimization is the preferred strategy of managing hazardous wastes, through elimination, substitution of less persistent or less hazardous alternatives, increased efficiency of industrial processes, containment and capture (for example, in chemical production, to reduce loss through leaks or venting), recycling, and re-use. Many laboratories and research centers around the world are working on safe and non-bio-persistent alternatives to traditional chemistry. Indeed, many chemical companies have identified “green chemistry” as a growth opportunity.

## **Special Wastes**

Medical, veterinary, biomedical research, and infectious wastes are prohibited from open dumping or transfer to landfills under RCRA unless they have been decontaminated from infectious risk. Almost all medical waste is incinerated under enclosed conditions in a rotary kiln. Incidents of dumping in the 1980s, in which medical waste washed up on shore on the East Coast of the United States, drew attention to medical waste as a management problem. However, it was not clear that this was a widespread problem.

The Low Level Radioactive Policy Act of 1980 mandates the Nuclear Regulatory Commission to oversee proper disposal of low-level radioactive waste, but this is ultimately the responsibility of states within their own borders. Since not every state has a facility, state compacts have formed to receive radioactive wastes from their participants, but there is still a shortage of facilities.

## **Superfund**

When an incident or the suspicion of trouble emerges with respect to a hazardous waste disposal site, the community may be

devastated. A particular concern is fear of cancer. This fear has effects on the community that may be far worse than the direct effects of the waste site. A major issue in dealing with incidents involving hazardous substances is the psychological effect of the incident upon the residents. The effect can be devastating and long-lasting. (See Chapter 24.) In the 1970s, the discovery or recognition of thousands of mostly abandoned toxic waste dumps in the United States led to congressional action to hold the responsible parties accountable for cleanup, or to create a mechanism for cleanup. The action also allowed for the expenditure of public funds, if necessary, if no responsible party could be identified (which has been the situation in 30 percent of cases so far).

In 1980, Congress passed the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), which provided a broad federal authority to respond directly to releases or threatened releases of hazardous substances that could endanger public health or the environment. It established prohibitions and requirements concerning closed and abandoned hazardous waste sites, and provided for liability of the persons responsible for the releases of hazardous waste at these sites. Cleanup of the remainder of the inventory or sites depends on state action.

CERCLA also created the Superfund program to clean up the most dangerous hazardous waste sites. Whenever possible, this was to be done without the use of federal money, by requiring action from the parties responsible. For those interventions that required federal funding, a tax was imposed on the petroleum and chemical industry. This financed a trust fund (the “Superfund”) for cleaning up abandoned or uncontrolled facilities when no responsible party could be identified. The original and reauthorized Superfund cleaned up only a small fraction of the nation’s hazardous waste site before the funding was exhausted in 2003, much of it expended on assessments and legal action rather than on remediation. Congress now funds the program by allocation from general revenues.

The Superfund Amendments and Reauthorization Act (SARA), which came in 1986, extended the mandate of the Superfund program,

fine-tuned the criteria, supplemented the funds, required that other federal and state regulations be taken into account for evaluation and remediation criteria, and coordinated efforts with state governments. SARA also had the important effect of shifting the emphasis from a “by the numbers” exercise of applying standards to a focus on human health risk, making Superfund more of a public health-oriented intervention. It also emphasized permanent solutions rather than stopgap control measures, and promoted innovative technologies, such as in situ treatment (destruction without removal from the site) and bio-remediation (use of plants). SARA also created a broad community right-to-know, which encouraged citizen participation, granted medical access to exposure information, and provided for emergency planning for industrial spills and accidents. Because of its role in risk perception, the community right-to-know provisions of SARA are described in Chapter 8.

Superfund is authorized to manage a list of designated sites called the National Priority List (NPL). These are sites that are evaluated and deemed to represent a sufficient hazard to health and the environment to require federal intervention. Superfund legislation provided for two kinds of response: short-term “removal,” which would respond to emergent or urgent situations; and long-term “remedial” response actions, which could be pursued when the situation was stable in the short term but posed a risk in the longer term. A decision to remediate then triggers engineering assessment and planning for safe construction, enhancement of destruction on site, or excavation as required. The objective of remediation is to clean up the site according to predetermined “remediation criteria,” based on existing standards and risk assessment, and to return the site to commercial or other use.

Registration with NPL is based on the score that a site receives on an assessment protocol known as the Hazard Ranking System, which takes into account all pathways of potential exposure. If the assessment suggests a need for emergency action, the site is first inspected to determine if there is an immediate danger to residents. The inspection may include testing to determine the chemicals present, and

whether they are migrating into the ground or water. To qualify for Superfund, a site must score enough points on the EPA's Hazard Ranking System. If it scores high enough, the site is entered in the National Priority List for federal attention. Thousands of sites around the country are known to be contaminated by hazardous waste, and tens of thousands have been evaluated by EPA. By the end of 1989, around the time that SARA took effect, only 38 had been cleaned up. That number has since increased greatly, and in 2007 the EPA reported about 2,000 sites "under control."

Many states have their own cleanup programs for sites that do not qualify under Superfund. These sites are less urgent and less extensive than those on the National Priority List. While most states have accomplished few if any cleanups to date, others have done a lot and have achieved considerable results. Most cleanups performed by states were done by New Jersey, California, and Massachusetts. These three states have what are by far the largest state programs now in effect. The ability of states to clean up major hazardous waste sites is variable, however. Most states have little experience in doing this, and their programs and budgets are small. Smaller and less affluent states would also have less clout in negotiating cleanups with the current site owners or responsible parties.

The EPA makes decisions about cleanup levels by applying various applicable environmental laws, air pollution standards, drinking water standards, and a hazards assessment of health risks proposed by the particular site. Cleanup levels are therefore set to reflect both health and environmental concerns.

Remediation is a technical process of achieving these cleanup levels through engineering and other methods. The techniques that are used include incineration, solidification, chemical treatment, evaporation, repeated flushing of soil, bio-degradation, containment, temporary storage, removal off-site, and pumping out groundwater.

The SARA amendments to the Superfund legislation require that preference be given to remedial actions that do not merely contain the hazardous waste, but also treat it to reduce its volume, toxicity, or mobility. Typically, remedies are chosen primarily to reduce the threat

of groundwater or soil contamination. Of course, the best approach of all is to reduce the generation of hazardous waste at the source, and current regulatory efforts are intended to do just that. For existing hazardous waste disposal sites, however, it is too late for this approach.

### Persistent Toxic Substances

“Persistent toxic substances” (PTSs) are environmental contaminants that persist in the environment (elements or organic compounds that have a long half-life; the EPA uses the criterion of eight weeks or more), that are bio-concentrated by species and bio-accumulate from one species and trophic level to another, and that have an effect on the environment. They include metals, such as mercury; organochlorine compounds, such as PCBs, dioxins, and furans; and many pesticides. PTSs by their nature tend to migrate and present management challenges that are not confined to single ecosystems or countries.

PTSs are a particular management concern in water. There have been several efforts around the world to manage PTSs cooperatively across borders. One of the most effective is the International Joint Commission (IJC), an intergovernmental deliberative body with representatives from Canada and the United States that was started in 1909. The IJC has been particularly active on issues involving contamination of the Great Lakes. It has also examined and made recommendations to federal, state, and provincial governments regarding rivers crossing the international border and trans-boundary air pollution. Another such international agency is the Baltic Sea Protection Convention (Helsinki Convention), administered by the Helsinki Commission (known as HELCOM). This commission is a pact of nine area states (with the recent addition of the European Union) formed in 1974 for the purpose of managing and ensuring sustainability of the Baltic, a shallow and highly threatened body of water.

In 1997, under the auspices of the United Nations, the Stockholm Convention on Persistent Organic Pollutants was introduced. Better

known as the “POPs Treaty,” this international agreement calls upon the signatory nations to monitor, reduce, discourage, and ultimately eliminate designated POPs compounds, including unused stockpiles, from the environment. The POPs targeted by the treaty are colloquially known as the “dirty dozen”: aldrin, chlordane, DDT, dieldrin, dioxins, endpin, furans, heptachlor, hexachlorobenzene, mirex, PCBs, and toxaphene. The POPs Treaty has been effective, and levels of POPs are declining throughout the world. It is considered one of the major success stories of environmental regulation on a global scale.

## **FOOD SAFETY**

Food safety is a ubiquitous public health concern. Food-borne diseases are generally familiar to the practicing physician. Suffice it to say that food service facilities in the workplace and employers’ facilities must comply with the same regulations as restaurants and public facilities. In addition to these considerations, the OEM physician may also encounter issues of food safety with respect to consumer protection and product safety, particularly in the food processing industry.

Many issues relate environmental health to public health, in both direct and indirect ways. Directly, issues of contamination and food quality are core public health issues. Although a minor issue in terms of risk, bovine spongiform encephalopathy (BSE, more familiarly called “mad cow disease”) demonstrated a novel pathogen (prions) transmitted to humans through the consumption of infected beef that produced a chronic neurological disease, the consequences of which devastated the international beef industry for a half decade. Indirectly, agriculture and food production are closely linked to OEM and public health through connections of nutrition, occupational health, pesticide residues, and food supply and security issues. Many emerging infections, often bird infections as in the case of avian influenza, can jump to human beings directly. Others, like SARS, require passage through animals, particularly swine (in the case of other forms of influenza), before they infect human beings.

Like SARS, they tend to occur in places where there is a dense concentration of agriculture and people in close proximity, as in China. The relationship between food and public health, therefore, is extensive. This section will focus on contamination of food, as this is the issue most likely to be encountered by the OEM physician.

Contamination of food may occur anywhere along the continuum of “fork to food” (a phrase common in the industry). The single most important factor is hygiene in the food handler, specifically hand washing, but there can be lapses and opportunities for contamination in many other ways.

Food-borne illness from microbial contamination is very common. Estimates (from the Institute of Medicine and the Centers for Disease Control) are as high as 76 million cases per year, with 325,000 hospitalizations and 5,000 deaths, even accounting for considerable underreporting, which biases estimates downward. Food-borne illnesses are easily misinterpreted or misdiagnosed as other problems, including influenza. Although these illnesses are usually minor and self-limited, they can be dangerous to immunocompromised people and patients with renal disease (a risk factor for listeriosis). Food-borne illnesses cause considerable discomfort, inconvenience, and lost work productivity. The problem is not confined to the stereotypical protein-rich foods kept at the wrong temperature, the iconic chicken salad of epidemiological training exercises. There have been recent, large outbreaks involving organisms not usually associated with the food in question: milk (norovirus, *Campylobacter*), unpasteurized apple cider (*E. coli* O157:H7), strawberries (hepatitis A), and processed meats (*Listeria*). This demonstrates that the pattern of food-borne outbreaks is changing as the high frequency continues.

Food-borne outbreaks may be caused by viruses, bacteria, parasites, prions (although very rare despite the attention to BSE), toxins, chemical contamination, and, rarely, radionuclides. A list of common contaminants is presented in Table 12.7. Virtually anything that can contaminate water can contaminate food, and vice versa. The range of potential contaminants is therefore enormous. Contamination may be deep (as in the case of *Salmonella* and eggs or fowl), or superficial (as

**Table 12.7.** Common and Noteworthy Food Pathogens

- 
- Bacteria
    - *Campylobacter* (most common)
    - *Staphylococci*
    - *Salmonella*
    - *Shigella*
    - *Listeria*
    - *Escherichia coli* (nontoxigenic, toxigenic, O157:H7)
  - Viruses
    - Hepatitis A (most common)
    - Norovirus (often explosive outbreaks, e.g., aboard cruise ships)
    - Many others (very common)
  - Parasites
    - Liver fluke (China)
    - Giardia (giardiasis): “beaver fever”
    - Trichinella (trichinosis)
    - Cryptosporidium
    - Cyclospora
  - Prions (variant Creutzfeldt-Jakob disease)—rare
- 

by fecal-oral transmission through poor hygiene by food handlers), or by spoilage (as by scambroid toxin affecting fish). Chemical contamination can occur through natural toxins (such as ciguatera), by biological contamination (mycotoxins), or by inadvertent chemical contamination (such as by atrazine, a pesticide). Chemicals that are added to food as additives sometimes have adverse effects (as with sulfite, which can trigger asthma in some people). In addition, there are many idiosyncratic and allergic reactions that affect individuals, such as peanut allergy (which seems to be growing in frequency and may be life threatening), and undisclosed gluten consumed by persons with celiac disease (which also seems to be increasing in frequency).

Food-borne illness can be minimized by thorough cooking, separation of foods likely to be contaminated (such as fowl, which should always be assumed to be contaminated with *Salmonella*),

**Table 12.8.** General Approach to the Investigation of Food-Borne Outbreaks

- 
- Epidemiological phase
    - Identify population at risk
    - Case definition
    - Calculate attack rates by risk factor
    - Collect any food that remains for culture, analysis
    - Intervention if possible
  - Laboratory phase
    - Recovery from food
    - Recovery from stool
      - Ova and parasites
      - Culture
    - Blood culture
    - Serologies (esp. hepatitis)
    - Serotyping and molecular techniques
- 

washing surfaces, and vector control. Microbial contamination is least likely in foods that are acidic, have a high osmotic load (although staphylococcus is resistant to the osmotic effect of salt), or have been heated. Freezing kills only some bacteria and does not kill spores. Cooling through refrigeration only slows bacterial growth but does not kill bacteria.

Some chemicals effectively preserve food and are used in the industry, but have other health risks: sulfite may trigger asthma in susceptible individuals, and nitrite forms nitrosamines in the body, which are carcinogens and need to be kept at low levels. Antibiotics are not an acceptable means of preserving food.

The general approach to investigating a suspected food-borne outbreak is outlined in Table 12.8. This task is the responsibility of public health officials and is rarely performed by OEM physicians.

The prevention of food-borne outbreaks rests on strategies at each step of food production, processing, and preparation. Table 12.9 outlines these strategies. In the 1980s, the traditional approach, which emphasized inspection by government inspectors, was perceived to

**Table 12.9.** General Strategies for Food Protection

Agriculture and production	USDA, EPA regulation
Food gathering and storage	Fumigation and contamination potential
Food processing	Contamination, additives
Food product distribution and storage	Contamination
Food preparation	Food safety guidelines, food service inspection

- Determine hazards and assess their severity and risks
- Identify critical control points
- Institute control measures
- Monitor critical control points
- Take action in event of noncompliance
- Verify that system is functioning

be largely failing and incapable of meeting the need in the face of increasing production. The Food and Drug Administration, in cooperation with the Food Safety and Quality Services of the U.S. Department of Agriculture, changed procedures about a decade ago, adopting a proactive approach based on corporate accountability and responsibility. The system is known as the Hazard Analysis and Critical Control Point System (HACCP). The purpose is to identify the critical points of likely contamination (cooking at insufficient temperature, for example) and to put into place measures appropriate to the problem (recording automated temperature monitoring and controls) to ensure that contamination cannot occur, and then to monitor the performance of those measures through surveillance and management tools (reviewing records and product testing). Despite the large number of food-borne outbreaks that have been reported in recent years, HACCP is credited with preventing many others and with allowing a more rational investment of resources for food protection.

Restaurant and food service inspection is a quintessential public health function that is the responsibility of local public health agencies. Food service inspectors follow a protocol that covers facilities (for example, ensuring that double sinks are available so that food preparation and dishwashing do not occur together). It also covers equipment (presence of a thermometer that is obviously in use), process measures (refrigeration temperature), housekeeping and the potential for cross-contamination, and behavior (handwashing, in particular). Inspectors are often seen as adversaries by proprietors, who generally operate on thin margins, because they can close down restaurants and other food services by order, or deny the service an "A" grade in jurisdictions that use this system, motivating queasy customers to go elsewhere. However, their services are indispensable and much appreciated by the small segment of the public that pays attention.

Other topics related to food are beyond the scope of this chapter.

## ECOSYSTEM HEALTH

Environmental problems can be divided into those that involve pollution and those that do not. The human health effects of pollution-related problems are usually more obvious because the same mechanisms of toxicity may affect human beings, and the same fate and disposition of chemicals in the environment that bring them to other species bring them to human beings as well.

Problems that are not directly related to pollution, such as loss of bio-diversity, resource depletion, or deforestation, are more subtle in their implications for human life. They may be associated with increasingly unstable ecosystems, changes in disease vector distribution, loss of economic opportunity, severe erosion, insecurity of food supply, impoverishment and its social consequences (which may include substance abuse, family violence, poor nutrition, and obesity), escalating commodity prices (especially for food and fuel), and socioeconomic changes that compromise social support, preventive services, and healthcare. The

health consequences of ecosystem disruption, while very real, are inordinately complicated and tend to be very specific to the situation. Most OEM physicians will not be faced with major responsibilities in this area (although some will). Ecosystem disruption will not, therefore, be discussed further in this chapter except as it relates to global climate and atmospheric change.

### **Ecotoxicology**

Ecology is the study of biological communities, with particular emphasis on the relationships among species, relationships between living organisms and the physical environment, and the flow of energy and nutrients among living systems. Ecologists concern themselves with species and their interaction, not individual organisms, and with systems rather than the behavior or physiology of a given species. The basic concepts of ecology, such as the “web of life,” “food chains” (trophic levels), carrying capacity, the population curve, and so forth, are beyond the scope of this chapter. They have in any case become part of the knowledge base of every scientifically educated person in this generation. Physicians have an advantage in understanding the behavior of chemicals in the environment even though they are not educated in ecology, because a similar process was part of their medical training. The various living systems resemble the human body. (There is even a school of thought in ecology, called the “Gaia hypothesis,” that takes this idea literally and considers the Earth as one enormous, integrated organism.)

“Ecotoxicology” is the fate and disposition of chemicals in the environment, relating to what they do to communities of species. Just as toxicology is fundamental to human health in the field of occupational and environmental medicine, ecotoxicology is fundamental to understanding the effects of chemical pollution on the environment. There are also many similarities and analogies to human toxicology that can help the physician to come up to speed quickly on problems related to toxic chemicals in the environment. The key is to think of

the environment as a complex organism, like the human body, with its own pathways of exposure, circulation, metabolism, and movement of chemicals and their products.

When substances are released into the environment, the phenomenon is similar to the idea of exposure and absorption in human toxicology. The contamination of air, water, and soil has similarities to the exposure pathways for human beings: inhalation, ingestion, and skin absorption. Airborne contamination spreads quickly, and behaves in accordance with the physical principles that govern transport and deposition. Water contamination is not unlike ingestion in that many of the principles governing the behavior of the contaminant relate to the chemistry and physical properties of the chemical; bio-transformation may take place due to the action of aquatic species. Soil contamination may lead, after migration and percolation downward, to entry into groundwater or surface area, but along the way there may also be bio-transformation. Soil fungi and bacteria function, in many ways, as the Earth equivalent of a bio-transforming organism such as the liver.

The functional equivalent of absorption in ecotoxicology is uptake by living organisms. Here the same principles apply. The lipid solubility of the contaminant plays a major role in determining which species accumulates the contaminant, and in bio-persistence and storage depots for lipid-soluble contaminants. Metals may bind to soil and clay, which may serve as a storage depot and result in sustained levels due to slow release, not unlike binding proteins for xenobiotics.

Distribution in human toxicokinetics has its counterpart in the circulation of chemicals and their migration to and from water, air, and soil. These pathways may become very complicated. For example, the migration of some lipophilic persistent organic pollutants, including chlorinated hydrocarbons, passes through phases in which the chemicals may be transported by air, absorbed and released from water, adsorbed and desorbed from soil or vegetation, and accumulated in species that are lipid-rich. The principles behind this process are the same as for the human body, apart from

important temperature effects governing the rate of desorption and release, which have the net result of concentrating these chemicals in colder, northern environments.

Bio-transformation takes place in the environment just as metabolism takes place in the human body. For the most part, the metabolic pathways are the same, or very similar. In the environment, the individual species in the biological community take the place conceptually of metabolizing organs, such as the liver, in which the chemical may be degraded, activated, or conjugated, just as in Phase I and II metabolism (see Chapter 2). Some transformative steps take place without bio-transformation, such as the degradation of many pollutants by exposure to ultraviolet light.

The one exception to this elaborate analogy between ecotoxicology and human toxicology is that there is no excretion pathway in the natural environment. The only pathway to elimination is by chemical transformation. However, chemicals may be sequestered for long periods in sediment or stored in reservoirs with essentially unlimited capacity relative to the amount ("sinks"), such as the ocean.

There are, of course, differences between toxicokinetics in the human body and the fate and disposition of pollutants in the environment. One of them is that in the environment, especially in water, organisms may take up, or "bio-accumulate," chemicals from one or many sources, and these organisms may "bio-concentrate" the chemicals, absorbing and keeping on board more than would be absorbed by passive diffusion alone. (This does not happen with pollutants that are not highly persistent, such as water-soluble organic compounds and chemicals that are readily bio-transformed.) When species are consumed, persistent pollutants will "bio-magnify," resulting in increasingly high concentrations up the food chain. The result, such as cadmium in the kidneys of northern caribou, can present a hazard to both wildlife species and human beings. However, bio-accumulation and bio-magnification are also important principles underlying bio-remediation, which takes advantage of these effects among plant species to remove persistent pollutants from bodies of water and soil.

The most important treaty governing persistent organic pollutants is the “POPs Treaty,” the Stockholm Convention on Persistent Organic Pollutants. The treaty is a United Nations initiative, ratified in 2000 and binding on 122 countries, that commits the parties not only to stop producing but also to eliminate and ban twelve particularly troublesome chemicals. The chemicals are the following: eight obsolete pesticides (aldrin, chlordane, DDT, dieldrin, endrin, heptachlor, mirex, toxaphene), hexachlorobenzene (an obsolete fungicide but mostly a contaminant in the manufacture of other chlorinated hydrocarbons), and three sets of structurally similar organochlorine compounds that are usually found together (PCBs, dioxins, and furans). None is economically essential in today’s world, and an exception is made for DDT for vector control if it is deemed essential for this purpose. (The use of DDT for malaria control involves much lower levels of application than the agricultural use that caused the ecological problems for which the substance was banned.) The treaty also provides for financial assistance to ease the transition to alternatives for countries that would face hardship, or for which it would not otherwise be economically feasible to collect and destroy remain stocks of the eight pesticides. The POPs Treaty is a landmark in international cooperation on chemical hazards. Although its direct implications for human health are limited, it is an important precedent as well as part of the regulatory framework affecting OEM physicians with responsibilities for environmental management.

### **Global Atmospheric Change**

Climate change is the most serious of a number of closely related atmospheric changes. “Greenhouse gases,” as most informed people know by now, are gases that allow infrared radiation to pass without absorption, but have an insulating effect, trapping heat. They are produced by most human activity, agricultural or industrial, with carbon dioxide emissions contributing the greater part. Other greenhouse gases include methane; chlorinated

and fluoridated hydrocarbons, which also play a role in stratospheric ozone depletion; and water vapor. The accumulation of greenhouse gases due to human activity has markedly enhanced the already strong “greenhouse effect” that maintains surface temperature on Earth. This has led to an accelerated increase in temperature over time, which is proceeding much faster than climate change models predicted. Although greenhouse gases are not themselves toxic, the ecological disturbances that they are causing may soon have devastating effects on climate, food supply, water availability, and health. Because heat retention involves the storage of energy, climate change has the net effect of making dynamic climate and ocean systems unstable, rather than causing a predictable or uniform “global warming.” The result is greater weather extremes.

A related problem is stratospheric ozone depletion, in which certain highly persistent chemicals scavenge and remove ozone in the stratosphere. Reduction in high-altitude ozone concentrations results in increased penetration of ultraviolet light to the surface of the Earth, because ozone absorbs UVA and UVB (see Chapter 9). For reasons that have to do with air stability and recirculation of the ozone-depleting chemicals, the major effect is seen in the atmosphere over the Antarctic, where depletion is most extreme, leading to the colloquial term “the ozone hole.” The result of more extensive ozone depletion would be markedly increased rates of skin cancers, including melanoma, and cataracts. Production of chlorofluorocarbons (CFCs), the related but less potent hydrochlorofluorocarbons (HCFCs), and individual chlorohydrocarbons (tetrachloromethane or carbon tetrachloride; 1,1,1-trichloroethane) is being progressively phased out as the substances are banned on a timetable and alternatives become available. International action coordinated by the Montreal Protocol on Substances That Deplete the Ozone Layer has dramatically reduced emissions of chemicals that catalyze the degradation of ozone. The Montreal Protocol is generally regarded to be the most successful international agreement on environmental issues ever concluded, and one of the most successful examples in history.

of international collaboration on a complex problem. A detailed discussion of these issues is beyond the scope of this book.

## ENVIRONMENTAL HEALTH AND CHILDREN

Environmental hazards steal away the future, creating risks that will have to be dealt with by the next generation, and limiting the safety and the choices of the next generation. In this sense, environmental hazards that are created in the world today threaten all children and cheat them of a future that might have been better. Because people are deeply concerned and protective toward children, particular attention has been paid to environmental health and children in recent years.

Children are physically more susceptible than adults to environmental hazards, for many reasons. Their bodies are still developing and the effect of an environmental insult can interfere with that development. Lead, for example, causes much greater damage to the central nervous system in children than in adults exposed to the same levels. The metabolic rate of children is much higher than that of an adult, in part because they are still developing, and in part because they are smaller in size. This means that their respiratory rate, for example, is proportionately greater, and they breathe in much more air pollution in relation to their body weight than an adult does in similar circumstances. The smaller size of their bodies means that the dose of contaminants is larger per unit of body weight, an important toxicological difference. If they are healthy, children are physically much more active than adults are, and they may come into contact with contaminants in soil while playing outdoors.

Children have a greater chance of experiencing chronic effects of exposure to environmental hazards because their lives lie ahead of them. If a child is exposed to a carcinogen, for example, chances are that he or she will live well beyond the latency period, the years that it takes for a cancer to develop after exposure. An adult, on the other hand, may not live long enough after such an exposure to get the disease, or may die of something else before the disease develops.

Children are also vulnerable in a social sense. The unique vulnerability of children arises particularly from social factors. Children cannot take care of themselves. They rely on a functioning society and infrastructure to protect them, and when society fails them, children suffer more than adults. Even in societies in which there is a functioning child welfare system that protects children, and a functioning public health system that looks after the health of its people, neglect, abuse, or ignorance at the level of the family may place the child at risk in his or her own community. When families are not strong and poverty is severe, children are often exploited and lose their rights. Children are less able to protect themselves by such means as making informed choices, protesting work conditions, and refusing to allow themselves to be exposed to hazards.

In some countries, children are forced to work in hazardous environments to earn money to support themselves and their families; under such conditions, which can amount to slavery, children have no protection and may be severely abused. There may be the expectation that children help support the family, resulting in long work hours at a young age with no opportunity to go to school. Children, even if mentally and physically capable of the work expected of them, may be emotionally immature, impulsive, and unable to understand their own inabilities to work safely. When they are injured, there may be no one to support them or to replace their lost income. Environmental hazards are threatening to all children, however, whether they are in such abusive situations or safe in a society that values and protects them.

Because children spend so much time at home and in school in developed countries, the problems of the built environment are central to children's environmental health. On the scale of neighborhoods, the design of communities, including the opportunities they provide for playing and walking safely, affects fitness and the risk of childhood obesity. This has been documented in the work of many investigators and brought to recognition by the work of Howard Frumkin and others. Passive smoking and exposure to indoor air pollution are well recognized as risk factors for respiratory disorders

and for low lung function in adulthood. More recently, dampness as a factor in itself, apart from mold growth and bio-aerosol generation, has been identified as a risk factor in children's respiratory health.

The potential exposures encountered in the built environment have been easier to identify and measure than the health effects that may be associated with them. The "hygiene hypothesis" has been central to studies and current thinking on asthma. This hypothesis suggests that children who come into contact with diverse antigens early in life are less likely to develop asthma than those who are protected during infancy and early childhood, and is supported by a large body of evidence. The hygiene hypothesis has cast doubt on conventional wisdom regarding housekeeping and health by implying that a house can be "too clean," and that early exposure in childhood to a robust mix of antigens may be healthy. Obviously, homes that are unsafe, dirty, have poor sanitation, and are host to disease vectors are not acceptable, and the hygiene hypothesis does not suggest that they are.

Conditions vary greatly among countries and are not always better in developed countries. In surveys conducted by WHO, developed countries have reported ambient (outdoor) air pollution and lead as the most common problems. In general, children's environmental health in most developed countries focuses on primary prevention of exposure to hazards unique or specific to children. In Europe and elsewhere, children's environmental health is understood to be a broad field in public health in which the emphasis is on primary prevention and effective public policy that protects children; the role of clinicians is more or less confined to counseling, documentation, and advocacy. The general approach is the forging of a "healthy public policy" in which proposed decisions at a political and local level are weighed as to their effect on children, and acted upon accordingly. Especially in Europe, children's environmental health protection is more closely connected to another environmental policy, that of the precautionary principle, which has been formalized as public policy since 2000. Children's environmental health protection,

as a policy, implicitly validates the “precautionary principle” because the public knows that effects across generations may be subtle and hard to predict.

The United States is unusual among developed countries because children’s environmental health has been dominated by issues involving lead. The United States was very late in banning lead paint (1978) in comparison to similar actions by other developed countries, which were taken in some cases as early as 1903. Consequently, the older housing stock in the United States is much more likely to present lead exposure risks from old paint than is the case elsewhere in the developed world. This problem has had a distorting effect on perceptions of environmental health, in part because the issue of lead and childhood risk has been mostly managed by secondary, not primary, prevention. In part because of the malign influence of lead as a hazard, in the United States pediatric environmental health has tended to emphasize diagnosis, screening, and tracking of exposed children more than in other countries.

In the United States, advocacy for children’s health uniquely transcends partisan politics and has played a central role in maintaining environmental progress in recent years. As a policy, children’s environmental health protection has sustained forward momentum in environmental protection regardless of the American political agenda or the opinions of elected officials. No politician or business leader wants to be seen as taking a position that injures children. The Office of Children’s Environmental Health at the EPA in the United States and the Centers for Disease Control, reflecting the vision of Richard Jackson, have been hugely influential in advancing children’s health protection at a time when other environmental initiatives have stalled and other federal government agencies have been reigned in. In addition, the usual way to block progress on an environmental initiative is to keep calling for more research, but the American public and decision-makers are less susceptible to this argument when it involves children.

In developing societies, children's health and the environment takes on a different profile. Developing countries have a wider range of common problems, including all of those in developed countries and more, such as childhood injuries, indoor air pollution (due to burning of bio-mass), infectious disease, poor sanitation with unsafe water, child labor, acute pesticide exposure, persistent organic pollutants, endocrine disruptors, and arsenic in drinking water. Risks to children reflect broader public health issues, such as poor access to clean water, inadequate sanitation, and air pollution (indoor and ambient), the latter of which contributes to high mortality from acute respiratory diseases in children. Climate change, as it is manifesting itself, as well as migration and internal displacement due to conflict and violence present particular risks to children. However, in most of the world, malaria, unsafe drinking water, and malnutrition remain among the greatest threats to children's health.

Economic development issues are inextricable from children's environmental health. Embedded in the context of broader social issues, such as lack of economic opportunity and low literacy rates, children's health and the environment issues seem more an effect than a cause, a consequence of development requiring broader economic solutions. However, unless the health of children is specifically addressed as its own agenda, the issues are likely to be marginalized by bigger budget issues, and hazards to health may inadvertently also be marginalized. Gas may replace charcoal as a fuel indoors, but if the home continues to have poor ventilation, an opportunity for greater health gain has been lost.

Children's environmental health is addressed by an increasingly specialized field of professional practice, called "pediatric environmental health" by physicians in the field of pediatrics in the United States. As a very broad generalization, American proponents of pediatric environmental health see it as a clinical specialty of pediatrics in which individual diagnosis and treatment are primary, prevention is on an individual basis, and the individual child is the center of attention. In Europe, on the other hand, the emphasis is

on collective action for the protection of children, a public health approach.

The principal international organizations for the protection of children from environmental hazards are agencies of the United Nations. The World Health Organization (WHO), an agency of the United Nations, compiles authoritative information on threats to children, monitors trends, and provides technical support and encouragement for programs at the national level. WHO works closely with the United Nations Children's Fund, universally known as UNICEF, which has a broader mandate for children's health and welfare. UNICEF has special programs on sanitation and clean water, disease prevention, and emergency response.

Numerous nongovernmental organizations support the goals of children's environmental health and elimination of exposure to environmental hazards, but relatively few are devoted specifically to children's environmental health. The most extensive is a network of individuals and organizations known as the International Network for Children's Health, Environment, and Safety (INCCHES). INCCHES relies on a network of member-correspondents for updated information and welcomes new participants. The European Union, the EPA Office of Children's Environmental Health, the National Institute of Environmental Health Sciences, and the Centers for Disease Control and Prevention all play a prominent role on the international stage.

In the United States, Canada, and some other countries there is a network of what are called "Pediatric Environmental Health Specialty Units" (PEHSUs), which are centers for education and community outreach on child health and the environment. PEHSUs are established in academic centers for public education, professional education, and response centers for public and parent inquiries on children's environmental health. In the United States, they are supported by the EPA and the Agency for Toxic Substances and Disease Registry, and are located in every federal region. The network is administered by the Association of Occupational and Environmental Clinics (AOEC). PEHSUs that are

locally supported in other countries, including Canada, Mexico, and Argentina, have joined the network.

## RESOURCES

- Cascio J, Woodside G, Mitchell P. *ISO 14000 Guide: The New International Environmental Management Standards*. New York: McGraw Hill; 1996.
- Yassi A, Kjellström T, de Kok T, Guidotti TL. *Basic Environmental Health*. New York: Oxford University Press; 2001.

## NOTEWORTHY READINGS

- Bingham E, Meader WV. Governmental regulation of environmental hazards in the 1990s. *Ann Rev Pub Health*. 1990;11:419-433.
- DeHart R. Accepting the environmental medicine challenge. *J Occup Med*. 1993;35(3):265-266.
- Ducatman AM. Occupational physicians and environmental medicine. *J Occup Med*. 1993;35(3):251-259.
- Goldsmith JR. *Environmental Epidemiology: Epidemiological Investigation of Community Environmental Health Problems*. Boca Raton, FL: CRC Press; 1986.
- Goldstein BD. Global issues in environmental medicine. *J Occup Med*. 1993;35(3):260-264.
- Joint Industrial Safety Council. *On the Relationship Between the Working Environment and the Natural Environment*. Stockholm: Joint Industrial Safety Council; 1996. [A brilliant synthesis in pamphlet format.]
- Lave LB, Upton AC. *Toxic Chemicals, Health, and the Environment*. Baltimore: The Johns Hopkins University Press; 1987. [Classic]
- McCally M, Cassel CK. Medical responsibility and global environmental change. *Ann Intern Med*. 1990;113:467-473.
- Mead PS, Slutsker L, Dietz V, McCaig LF, Breese JS, Shapiro C, Griffin PM, Tauxe RV. Food-related illness and death in the United States. *Emerg Inf Dis*. 1999;5(5):608-625.
- National Academy of Sciences. *Global Environmental Health in the 21st Century: From Regulation to Corporate Responsibility*. Report of the Workshop. Washington DC: National Academies Press; 2007.

*This page intentionally left blank*

# **13 STRESS AND PSYCHOSOCIAL FACTORS AT WORK**

Stress has emerged as a highly visible occupational risk in modern working life. This does not mean that work was not stressful in the past or in other economic regimes. Work in the contemporary technological, and service economy might even be less stressful than in years past, when the consequences of failure or loss of work meant profound deprivation and even starvation for workers and their families, and workplace hazards abounded. This is still the case in most developing economies. However, there is a strong sense among the public and among occupational health professionals that modern-day stress is a new problem. It has certainly assumed a higher priority in recent decades.

The emergence of stress as an occupational health priority is in part a reflection of the assumed (but arguable) progress that has been made in eliminating the physical, chemical, and biological hazards, and in part a reflection of the accelerating pace of daily life in modern, technological societies. The increasing concentration of workers in the service sector and in jobs that are not physical has thrown a spotlight on other adverse working conditions, especially workplace stressors.

It may be that what is new about stress in the modern workplace is not a matter of degree but an absence of focus, which makes anxiety hard to cope with, and an absence of resolution, which prolongs the stressful experience beyond what people are physiologically programmed to endure. Workplace stress in the contemporary economy appears to be linked to object-less, free-floating anxiety and the de-personalization of the workers' experience, called "anomie" in the sociological literature. The concept of anomie is that workers have no stake in the production process other than the security of their wages, and therefore they can take little pride in the outcome or the quality of their work. This leads, especially in large organizations, to a feeling of disempowerment, a lack of identification with the employer or the work, and declining productivity. There are three obvious ways to address anomie: as a problem of productivity requiring a program of intervention among workers, as a problem of individual attitude that is detrimental to the worker as well as the employer, and as an organization problem requiring cultural change at the level of the employer. The theory of anomie has given way in recent decades to the "demand-control" model, described below, as the dominant model for understanding workplace stress, but it remains a powerful social explanation.

In their various roles as citizen, parent, manager, worker, and individual, all people in modern, urban life are under some degree of stress. Job-related stress is one of many sources of stress in daily life; other sources of stress may include family, financial, and personal concerns. One may imagine that stress from these various sources is additive. However, it is more likely to be interactive because stress in one area of human life tends to interfere with one's ability to cope with stress in another.

Job-related stress has many components, reflecting the content of work and the context in which the work takes place. Stress may be generated by many and various factors, including the pace of work, the control one exerts over the task at hand, the control one exerts over the immediate workplace environment, time flexibility, shift work, the respect from others one perceives, interpersonal conflict,

bullying and other forms of harassment, the risk of violence in some workplaces, and many others.

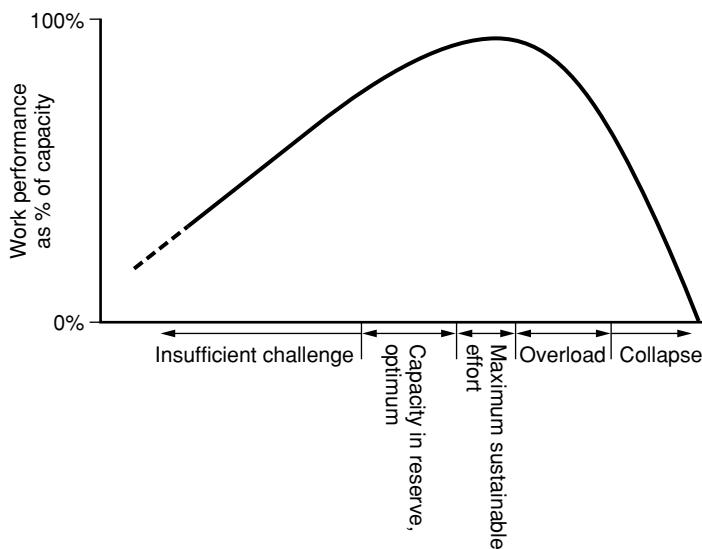
Stress investigators have developed the concept of “allostasis,” the condition of responding to a stressor or stressors beyond comfortable or homeostatic adaptation, and “allostatic load,” which represents the cumulative effect of adaptive responses to stress in excess of the ability to cope. Repetitive episodes of severe stress punctuating long periods of boring tedium where vigilance is required appear to be especially stressful. This pattern of stress is familiar from the work patterns of firefighters, hospital emergency rooms, nuclear power plant operators, and refinery operators. Such jobs are increasingly common as automation reduces the need for close human supervision of industrial processes.

Factors inducing stress, called stressors, may range from the profound, such as a deep personal or ethical disagreement with the goals or conduct of the organization, to the seemingly trivial, such as continuing interruptions or the constant, unrelenting annoyance of a workplace that is ill equipped or poorly designed. Some stressors are physical or physiological in nature, such as extreme cold, blood loss, severe illness, and loud noise. In other words, a stressor is what induces stress and provokes the response to stress, rather than a material category of hazard such as chemical hazards. What constitutes a stressor may vary from individual to individual as a consequence of learned experience, cognitive perception, physiological thresholds, phobias, and contemporaneous life events. For example, the prospect of being laid off from work may provoke little stress in a young, single person with marketable skills and an episodic work history, but may be extremely threatening for older workers who support their families, are worried about security during retirement, and doubt their ability to retrain. Consequently, the magnitude, cumulative dose, and direction of stressors have proven extremely hard to measure.

The effects of psychological stress in the workplace on human health are also difficult to quantify. The outcomes associated with stress may be physical or behavioral, and are complex, difficult to quantify, and subject to modification by social support and individual

host factors, collectively called resilience. Indeed, there are many positive forms of stress that are associated with enjoyment, constructive excitement, prowess, achievement, and motivation. The effect of stress on an individual depends in large part on that individual's personality, cognitive framework, coping mechanisms, social network, confidence in the future, and emotional makeup. Collectively, these factors, and others, contribute to resilience.

Constructive stress is motivating and energizing. The arousal-performance curve (see Figure 13.1) describes how increasing levels of stress increase individual performance—for example, work productivity—up to a peak. Beyond this peak, however, stress becomes destructive, and performance falls; it typically declines first in quality and later in quantity. Constructive stress, which is also called “eustress” or “positive stress,” occurs when stress is perceived as having a purpose and is followed by a sense of release or satisfaction. Excitement, a sense of commitment, teamwork under pressure,



**Figure 13.1.** The arousal-performance curve. Increasing levels of stress are associated with improved performance up to a peak. Performance declines after the point of overload, which varies from individual to individual.

achievement, and successful response to a disaster can all be exhilarating, energizing, and constructive emotions. Positive stress is a highly individualized response, however. It cannot be imposed on an unwilling worker by demands of management. It does not carry on indefinitely, but rather ends with a sense of accomplishment. It addresses clear goals the worker can identify with, not the personal goals of a supervisor. It is associated with team building and camaraderie, not divisiveness, and not with wasting time or waiting for hours or days for something to happen. Positive stress implies a goal worthy of the worker's efforts. Positive stress is even sought after. Risk-taking behavior, both in personal life and business, is a common attribute of high-performing chief executive officers and entrepreneurs. The excitement associated with winning sports teams and active vacations is highly constructive, in part because it is voluntary on the part of the individuals engaged.

Destructive psychosocial stressors in the workplace lead to a state of destructive stress, sometimes called "distress" or "negative stress," which has adverse consequences for health, emotional satisfaction, social adjustment, interpersonal relations, and productivity. Stress in this context may be considered as an adverse physiological response induced by the psychological reaction to an imbalance or mismatch between demand and capacity, expectation and comfort level, capability and demands of the job. Most recently, attention has focused on the role of individual factors that allow the worker to cope, collectively called "resilience."

The study of stress has been considerably advanced by the introduction of the "demand-control" model by Robert Karasek in 1979. This model evaluated the magnitude of workplace stressors in two dimensions: the demand or pace of work, and the degree of control that the worker has over the work environment. High-demand jobs with low degrees of control and decision-making were found to predict risk for myocardial infarction, a finding repeated many times. The demand-control model is now considered the core theory of occupational stress. Other models, however, seem to be complementary to the demand-control model and provide independent explanatory power.

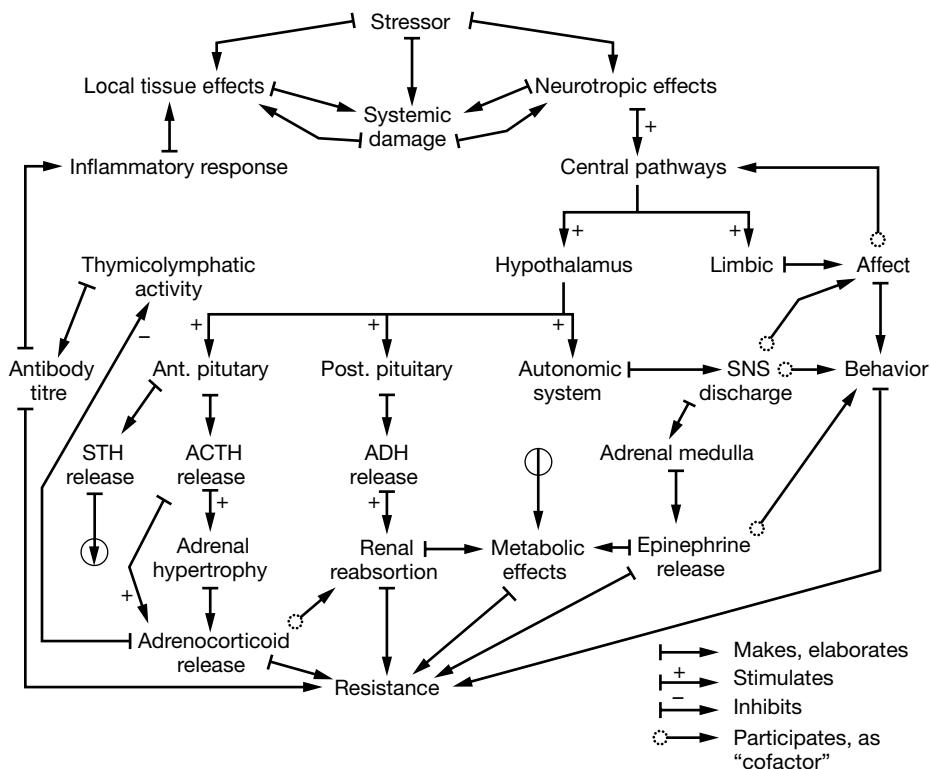
Another highly productive theory is the “effort-reward” model developed by Johannes Siegrist. This model and the instruments based on it attempt to assess the effort put into responding to work demands, including engagement and motivation, against the level of reward that returns to the worker, including material rewards (pay and employee benefits), professional opportunity, and intangible social benefits (team identification, camaraderie, social support, and sense of mission). The effort-reward model appears in many studies to add power to the demand-control model. Michael Marmot and colleagues, especially in the Whitehall study of British civil servants, have identified an independent contribution of social status and rigid hierarchy as workplace stressors. Another model in common use is the “occupational justice” model, developed by Ann Wilcock, which deals with how fair the worker perceives work to be, including the behavior of his or her immediate supervisor, the process of work organization, and dispute resolution. These models provide a framework for assessment and are the basis for instruments, mostly paper or Web-based questionnaires, that quantify exposure to psychosocial stressors.

The physiological pattern of the body’s response to stress can be divided into two general stages. These stages have been demonstrated in animal studies, but although they are generally applicable to humans, the additional factors of human cognition, culture, and learned behavior modify the effects of the basic biological responses. For this reason, among others, a direct extrapolation cannot be made between findings in animal studies and human individuals and populations. This was demonstrated years ago when a series of studies conducted by John B. Calhoun on crowded populations of mice suggested catastrophic consequences that might be seen in urban human populations. This impression was reinforced by the urban crisis of crime, decay, and inner city poverty that was emerging in the United States at the same time. Many opinion leaders concluded at the time that efforts at reform were pointless because the adverse response to crowding appeared to be innate. Subsequently, social scientists recognized that this pessimistic conclusion did not take into account the structure of social interactions, the context of crowding

(whether voluntary and familiar or imposed), and the differences between rodents and primates.

The first stage of the physiological response to stress is the well-known “flight-or-fight” response, mediated by the sympathetic nervous system and circulating epinephrine. The response was first described by physiologist Walter Cannon in the 1920s. By heightening awareness, mobilizing glucose stores, diverting blood away from internal organs and to muscle, the flight-or-fight response prepares the individual for short-term exertion in order to flee or to attack what is perceived as an imminent threat. This reaction occurs in response to physical stressors (such as extreme cold) as well as psychosocial stressors. The acute stress response was adaptive evolutionarily, and is still applicable when human beings face physical aggression by others or concrete physical threats, and must either deal with or escape them. However, when the threat is nonmaterial, the stress response becomes maladaptive. In the modern workplace, threats are mostly perceived rather than real; they are nonmaterial, sustained, and have no satisfying emotional resolution. Repeated many times during the day in response to constant or repetitive threats, perceived or real, the flight-or-fight response causes dysfunctional stress-related symptoms, and ultimately physical and behavioral problems.

The second physiological stage of stress was described by Hans Selye, an endocrinologist. He discovered that the response to prolonged stress is mediated by cortisol release from the adrenal cortex. This stage he called the “General Adaptation Syndrome” as it applies to rodent studies, although the comparability of this stage of response in rodents to that in humans has been questioned. Modern research has shown that while elements of the General Adaptation Syndrome certainly exist in humans, the integrated response is modified by cognitive behavior and perception. In rodents, a prolonged state of extreme stress leads to adrenal insufficiency, immune deficiency, changes in brain morphology (atrophy of hippocampal circuits), and eventual exhaustion and collapse. In human beings, there may be mild blunting of the immune response (the classic example being the frequency with which students get viral infections around



**Figure 13.2.** The General Adaptation Syndrome, mediated by cortisol, as elaborated by Selye and associates.

examination time), but the principal responses seem to be cognitive, not hormonal. Figure 13.2 illustrates the pathophysiology of the chronic stress response as elaborated by Selye and his school of thought.

The second stage of stress in human beings is associated with sleep disorders, eating disorders, weight gain, panic attacks, depression, and chronic fatigue. Affected individuals may unwise resort to self-treatment through alcohol or drug abuse, dysfunctional sexual behavior, gambling, or other maladaptive responses to anxiety. The risk of suicide and impulsive, self-destructive acts is dramatically increased. Recently, new evidence has suggested that poor dietary choices and metabolic changes favoring obesity may result from prolonged psychogenic stress.

Behavioral responses to stress may be somewhat crudely categorized as subjective, objective, and clinical. The subjective response to stress may be felt only by the individual worker and may not necessarily be evident to an observer. This may include anxiety, emotional lability, anger, depression, attention deficit (here meaning simply inability to focus for a sustained period), diminished short-term memory, and obsessive thoughts. Although not usually obvious to others, the subjective responses may be accompanied by manifestations of hyperventilation and sympathetic autonomic hyperactivity, such as sweaty palms, racing heart rate, dry mouth, and a slight tremor. The objectively observable response (here meaning visible to others) is usually seen as a change from the individual's previous personality or behavior, and may include irritability, reduced productivity, distraction, withdrawal from the social environment, impulsive behavior, family discord, diminished (or occasionally increased) sexual drive, disengagement from the events surrounding one, and symptom magnification (especially degree of pain). The subjective and objective responses may fall within the wide range of "normal" observed behavior, and therefore evidence of change is more compelling than any one feature. Clinical responses here mean those resulting in clearly abnormal outcomes, and may include neurotic adjustment disorders, affective disorders (including so-called reactive depression), anxiety disorders (including, if severe, posttraumatic stress disorder), diminished sexual performance (which is very disturbing to most people and becomes a secondary source of stress), somatoform disorders, sleep disorders, hypertension, and substance abuse. Some manifestations of these clinical disorders are undoubtedly secondary, arising from sleep disorder.

### **Individual Factors**

Stress feeds on itself: people can become neurotically stressed by worrying about the presence of stress in their life. Because mood, sleep, and other behavioral patterns are altered by stress, stress itself reduces one's ability to cope with the cause of the stress as well as

other hazards and stressors in daily life and work because of distraction, rumination over the problem, and the grieving process; the stress may even affect immune function. Stress complicates its own resolution with the sleep disturbance, depressed affect, feelings of inadequacy, and anxiety that it provokes.

As a risk factor, stress seems to act most strongly on pre-existing health conditions. There is a tendency to think of stress as one might think of other occupational hazards, assuming that exposure to a stressor results in an increased risk of a predictable outcome or a cluster of outcomes that can be monitored much like injuries or occupational diseases. However, stress is different from other occupational hazards in several important respects. There is no one, specific outcome associated with stress, as illustrated by Table 13.1. Psychosocial stress often magnifies the effects of physical stressors and heightens perception of the severity of the discomfort and disability associated with many chronic pain symptoms and illnesses.

**Table 13.1.** Conditions with a Strong Association with Stress

Headache	Depression
Low back pain	Insomnia or hypersomnolence
Gastrointestinal symptoms	Irritability and emotional lability
Spastic colon	Alcohol abuse
Duodenal ulcer	Decreased concentration, distractibility
Chest pain	Eating disorders
Labile hypertension	Sexual dysfunction
Autonomic effects (e.g., clammy skin, sweating palms)	Violence
Panic attacks	Emotional lability
Control of diabetes	Posttraumatic stress syndrome
	Change in personality

N.B. Note the absence of asthma, which for many years was wrongly attributed to stress.

Stress at work is intimately mingled with stress in one's personal life; just as the troubled and distressed worker brings some job stress home, the worker brings some degree of home life stress to the job. Background stress invariably exists at home and in the worker's personal life. If one's family life is unsettled, then the same degree of job stress may act more intensively on the worker because he or she is less able to cope with it at this time.

Mental health can be considered in terms of personal competence and ability to cope with social realities. Personal competence implies a sense of mastery and control and, particularly, resilience, the latter of which is defined as the capacity for coping with life situations. People usually deal with stress by resolving the source, avoiding it, sublimating or displacing their response, or ignoring it. People whose stress levels are resulting in distress and dysfunctional behavior patterns need help in resolving their response to stress. Although professional intervention is sometimes necessary, the first line of support for coping with stress is the worker's family and social network. Social support, which is experienced as resilience, is provided through social or helping networks: family and spouse, neighbors, friends, co-workers, fraternal organizations, alumni associations, bar patrons (with obvious drawbacks), and whatever other social interaction has meaning for the individual. People with few sources of social support are more vulnerable to stress than those who have the problem-solving and supportive resources of a natural helping network to draw upon.

Clinical mental health problems are a special case. Individuals with premorbid psychiatric diagnoses, such as anxiety disorder, depression, or panic disorder, may be especially susceptible to workplace stress. Accurate diagnosis and effective management of the underlying condition is the most important management approach to such individuals. Stress management and resilience skills may be useful, but they will not get to the root cause of the problem. Individuals with mental health problems require identification and referral to qualified practitioners. Employee assistance programs (see Chapter 19) are the appropriate means to address this.

Workers who are in jobs they are ill suited for, either because they chose their careers unwisely or because the job was the best or the only one available at the time, often feel dissatisfied even if they are successful at the job. These people may continually feel that there is something missing in their lives, and unless they realize their predicament, they become more and more dissatisfied and may experience burnout.

Women with both careers and families are a special group at risk for stress in the workplace. They actually work at two jobs, at home and in the plant or office; therefore, they have double the level of responsibility and work, but get only a single paycheck. They must fulfill the needs of others both at home and at the workplace while trying to fulfill their own needs as a person. As the traditional caregivers, they may have to balance a job that supplies needed income, child care, and care of a dependent parent or loved one, all at the same time. (Those who are caught between the need to provide care to an older generation and the responsibility of raising a younger one are often called the "sandwich generation.") Some cannot maintain the balancing act without paying a high price in stress.

Family stress comes from many life issues, but some of the more common ones are impending marriage, birth and infant care, troubled adolescent children, care of aging or disabled parents or other relatives, debt, divorce, death or disability of a loved one, impending retirement, and serious illness. Each requires the worker to establish priorities, set goals, find sources of support, manage change, and get knowledgeable help in identifying options and strategies for dealing with the problem. This is much easier said than done. Employee assistance programs (see Chapter 19), which are available to help with some problems, are not usually designed for these common stresses of adult living.

More individuals from disadvantaged minority groups are assuming leadership roles in business. Not only does this bring the increased stress associated with a position of higher authority, but these individuals now often find that they must communicate on many levels, maintaining relations with their peers and friends while at the same time becoming accepted by their subordinates. Added to

this may be the anxiety associated with the fear of failure in the new position, sometimes rooted in old thought patterns.

## **Work Organization**

Combinations of factors within and outside the work situation interact and contribute to disease; therefore, the study of the relationship between job stress and illness is complicated. A growing body of knowledge documents the effects of occupational stress in the etiology and development of physical and mental disease. Employers need to be aware of the sources of work-induced stresses that may affect their employees and their operations. The ideal organizational culture is one in which the worker feels free to express concerns and to ask for help.

One of the most common situations associated with stress in occupational settings is excessive demand for increased production. Working overtime or for long hours can be sustained for short periods, but if it continues, eventually fatigue, frustration, and ultimately stress-related symptoms set in. However, there are many other causes of job-related stress, as summarized in Table 13.2. Many investigators believe that the frustration associated with boredom, inactivity, and a sense of wasting time can be as stressful as overload, time pressure, and interpersonal conflict, particularly for those whose self-esteem is heavily bound to their work.

Stress occurs at all levels of employment. Supervisors have bosses over them, and managers at all levels are evaluated by superiors on a higher level. The chief executive officer is evaluated by a board of directors, which in turn is evaluated by shareholders. To minimize levels of stress, interpersonal relationships on all levels should be satisfactory, but need not be perfect. In the real world, such harmony does not occur, and the greatest stress tends to be concentrated at the points in the organization where demands are greatest but control is least.

If the work environment is a significant contributor to stress, management may consider changes in work organization that improve the quality of working life. These may involve increased

**Table 13.2.** Some Factors Contributing to Stress in the Workplace

*Work Organization*

- Job insecurity (risk of termination of employment)
- Isolation
- Overload
- Lack of control
- Shift work
- Paced work, speedup
- Company mergers, acquisitions, and bankruptcies

*Career Factors*

- Job insecurity (working on contract, part-time, temporary work)
- Stagnation
- “Career arrest,” failure to advance or “hitting the glass ceiling”
- Lack of challenge
- Inadequate credentials or qualifications (as perceived by the worker)
- Inadequate training or preparation to perform job

*Social Issues Internal to the Workplace*

- Personality clashes
- Role conflicts
- Unclear job description
- Poor communication of essential information, including goals, from superiors
- Poor communication upward of actionable information from subordinates
- Ambiguous or contradictory directives
- Ambiguous or contradictory performance expectations
- Unfair evaluations or labor practices
- Lack of recognition for performance
- Discrimination
- Harassment (sexual harassment, bullying)
- Cliques and social exclusion
- Disagreement with management policy

*Social Issues External to the Workplace*

- Public disapproval (of the business, company, job, or worker’s specific role)
- Family disapproval (of the business, company, job, or worker’s specific role)
- Declining reputation of employer
- Economic conditions affecting employer
- Economic conditions affecting personal finances

*(Continued)*

**Table 13.2.** (Continued)

---

<i>Job Characteristics</i>
Absence of personal identification with work
Employment in a job substantially below abilities
Promotion to a position that subjectively feels inappropriate with respect to relationship with peers, social class, age, or level of education (often called “status inconsistency”)
Promotion to a position in which one supervises former colleagues or friends
High responsibility/low authority to manage
Social isolation (depending on personality)
<i>Personal Issues</i>
Frequent contact with “stress carriers” (individuals who demonstrate anxiety-inducing behaviors: highly anxious, indecisiveness, demanding and “workaholic,” poorly skilled socially)
Concealment of alternative behavior or lifestyles perceived as “deviant” by the majority (i.e., being “in the closet”)
Unresolved emotions associated with fellow workers or supervisors
Family stresses, especially those that interfere with one’s ability to deal with workplace stresses (e.g., by causing sleep disturbance or caring for a family member requiring frequent medical appointments)
Phobias and other major mental health issues
<i>Physical Factors</i>
Unpleasant working environment (temperature out of comfortable range, excessive noise, visual interference, poor illumination, etc.)
Inefficient working environment (as perceived by worker)
Long or unpleasant commutes

---

worker participation in the decision-making process about particular jobs, plans, or operational decisions. A higher level of participation at the workplace level may increase satisfaction, stimulate interest, and result in higher productivity. Other approaches that might be tried include changes in reporting structures, changes in the division of labor, changes in control and monitoring procedures, increasing control by the individual worker over the process or job he or she performs, and changes in the social interactions among workers. If there are power struggles in the workplace, leaderless or self-organized

work teams may help to reduce perceived competition and personal alienation. Managers can share problems with their subordinates as a team, rather than attempt to conceal them. Workers often have well thought-out solutions to workplace problems, but may feel constrained from offering them to management.

In the 1920s, a series of experiments was conducted at the Hawthorne assembly plant of Western Electric, near Chicago. The experiments involved adjusting levels of illumination, work pacing, and timing of breaks and were carried out over several years. The results of these studies, particularly the one involving illumination, were interpreted and widely promoted as suggesting that any benign intervention, such as either lowering or raising illumination levels, increased productivity as a result of boosting morale among the workers through a sense of receiving attention and recognition for their contribution. However, these studies were incompletely documented at the time. In recent years, many scholars (led by H. McIlvaine Parsons) have examined the original, incomplete records of the study and interviewed the now very aged surviving participants and investigators. They concluded that this finding cannot be supported by the actual data, and that the conditions of the experiments were not accurately described at the time. There is also fragmentary evidence that production in the key experiment, involving a very small number of relay assemblers, did not rise. Therefore, the "Hawthorne effect," by which worker productivity increases non-specifically as a result of attention regardless of the actual intervention, appears to be a myth. Like many myths, however, it may hold a grain of truth in that while attention may not guarantee improved performance, neglect usually results in decreased performance.

Work-related stress is also associated with status in the hierarchy, but in a complicated way. Marmot and colleagues have demonstrated, in their landmark study of British civil servants, the Whitehall Study, that a wide range of adverse health outcomes is associated with low status. Other studies have confirmed the finding that, after adjusting for other health risk factors, people with lower social status are at higher risk of death, cardiovascular disease, and absence from work.

However, it is unclear that status works apart from the demand-control model or material privation, contributing independently to risk of ill health. The studies reported to date have generally been done on highly status-conscious populations, rather than on societies with fluid social mobility.

Other studies suggest that social status cannot be treated as a single variable. For example, an important Scottish study suggested a cumulative effect of status on mortality of only 3 percent for men who had already demonstrated some social mobility but remained in predominantly low socioeconomic status. However, an effect of 27 percent was suggested for subjects who were permanently stuck in lower socioeconomic status. This raises the possibility that opportunities for social movement are themselves dramatically protective, possibly as important as the status attained. On the other hand, social mobility may be just another marker for favorable health risk, such as prior good health. People who have the energy and drive to move up the ladder are probably different from those who fail to do so or who go down the ladder. One gap in these studies of status and health is that they do not take into account those who choose an alternative ladder, one that may lie outside the mainstream. Status outside of work, in the community or among peers, is not well studied because it is harder to measure.

Perhaps the single greatest work-related stressor is unemployment. The loss of a job is closely associated with subsequent mortality and ill health.

## **Shift Work**

The essential challenge of shift work has always been to obtain the economic or practical benefits without compromising the worker's ability to do the job and live a satisfying personal life. However, the problem took on new urgency in 2007 with the declaration that shift work may also place workers at risk for cancer.

In 2007 the International Agency for Research on Cancer concluded that there is limited evidence (2A in their system) for an

association between shift work and cancer in humans. The human organism follows a circadian cycle that is reset by exposure to bright light, and that reduces alertness and productivity in the early morning hours. The neurohormonal basis of adjustment to shift work involves synchronizing release of melatonin to light/dark cycles. Interruptions in light exposure or changes in shift may cause periods of low circulating melatonin levels. Because melatonin, which induces or maintains sleep, inhibits cancer, there has been speculation that a deficiency of melatonin may promote cancer. Whether or not this is in fact the underlying biological basis, a number of well-designed studies of shift workers and night workers show an association between shift work and cancer risk. The studies have mostly been done on nurses and flight attendants with regard to the risk of breast cancer. This is in keeping with animal experiments that show increased tumor rates associated with disturbed sleep-activity cycles, constant light, light at night, and simulated jet lag induced by changes in lighting periodicity. This has introduced a new concern to a field previously occupied primarily with issues of adjustment and neurocognitive functioning.

The popularity of shift work has increased over the last decades, especially with increased expectations that businesses will provide 24-hour service. The increase in shift work appears to be due to two major factors. In manufacturing, shift work exists primarily for efficiency in plant production. Shift patterns in manufacturing are often related to the need to keep plants or expensive equipment operating. Technological advances, energy costs, and more efficient usage of heavy machinery have resulted in a preference for continuous production in order to reduce costs of equipment maintenance and shut-down. It is often more profitable to run a plant “round the clock” in order to raise the level of annual operating hours and the profit. Productivity during some shifts may actually be rather low with respect to worker input, but equipment productivity is usually more critical in these industries. Machines work longer hours than do their human counterparts. Shift work not only saves capital, but reduces operating costs as well. Shift work may also be instituted to

allow production at times of reduced cost, avoiding peak load electrical rates, and to decrease traffic congestion.

Patterns of shift work in service industries tend to be more of a function of necessity or custom, rather than of efficiency. Where there is a need or consumer demand for certain services on a 24-hour basis, shift work is unavoidable. This has long been the pattern for medical care and public services such as fire and police protection and utility services. With increasing emphasis on personal convenience, retail, food, and just-in-time delivery, consumer services and transportation are increasingly available throughout the day, requiring staffing by shifts.

Despite the constant performance characteristics of plant equipment and the ongoing demands of society, human performance and reliability to meet these demands do not remain constant. Significant reductions in performance occur on a night shift, as measured by reaction time, accuracy of perception, and error rate. The nadir of performance appears to occur between 3:00 and 6:00 a.m., corresponding to the lowest point in the diurnal variation of body temperature, and is further reduced by the number of days spent on the night shift in a particular week.

Jobs requiring repetition, mental acuity, and vigilance regarding signs or observing details do not lend themselves as readily to satisfactory shift work as physical labor. Visual perception and response times are prolonged at night, and in conjunction with fatigue and sleep deprivation. Unfortunately, jobs requiring these higher-level skills include many occupations that are assigned by shifts, such as nursing and transportation workers, including airline pilots. Workers on a permanent night shift or a rotating shift appear to be affected mainly by the disruption of circadian rhythms and reduced contact with their family and the community. Physiologically, the body has many circadian rhythms: body temperature, pulse, breathing rate, chemical composition of the blood and urine, and brain activity. These rhythms are usually well synchronized by external events, such as light/dark cycles and the alarm clock, but can be artificially altered by long periods of isolation from external stimuli. Permanent night

shift workers show decreased amplitude as well as desynchronization of these circadian rhythms.

The periodicity of the human circadian rhythm is approximately 25 hours; it is reset daily by routine and visual cues, including daylight. It is easier for humans to postpone sleep than to advance the time of sleeping or get up earlier. Shift rotations of day-evening-night, either having a day off or working a double shift on the day of transition, are therefore better tolerated in general than day-night-evening rotation. It takes approximately fifteen days to adjust fully to a new shift, approximately one day for each hour of change in one's bedtime. More rapid schedules of rotation, while acceptable to many workers, do not permit synchronization to stabilize at any point, and are disliked by most workers.

The diurnal (day-night) cycle of body temperature appears to be one of the cycles most sensitive to change. Studies have shown a correlation between body temperature and performance, as judged by vigilance, accident risk, and manual dexterity. The optimum time to work is when the body temperature is high during the day. When these cycles are changed, for example by a transcontinental flight, the change in the pattern of body temperature seems to determine levels of performance, rather than the actual time of day.

Night workers typically get less sleep than day workers, and the sleep they do get is often of poor quality. Rapid eye movement (REM) sleep, which occurs more frequently after a person has been asleep for more than five or six hours, is a biological necessity. If a person is deprived of REM sleep, more of it will appear earlier on subsequent nights. Night workers who sleep at odd hours may use hypnotics to aid in falling asleep, and their sleep may be disturbed by the medication. They may not get adequate REM sleep and may develop chronic fatigue.

Of particular concern is the increased risk of motor vehicle incidents (especially during the commute home from work) and events of injury on the job among workers on night shifts. At least some of these events are due to lapses in attention and falling asleep behind the wheel, but fatigue-related impairments in perception and

judgment also play a role. Sleep disturbance plays a major role, especially in rotating shifts.

A few workers like shift work, some strongly dislike it and never adjust, and many have simply learned to live with it. Permanent shift workers may be different from other workers in matters of job preference, personality, or social ties; therefore, there may be a selection factor that favors certain individuals for night work. Aging workers tend to have more trouble adapting to shift work. Night shift workers tend to use sleeping pills more often than day workers do, and there may be other differences in use of medication and alcohol.

There are some obvious advantages and disadvantages to the various shift schedules. Steady day shifts have the most advantages as well as fewer disadvantages. A day shift schedule fits best into the family's schedule as well as social schedules. The disadvantages, which are relatively minor, include the worker's inability to tend to outside personal business during the day, and the fact that the worker cannot sleep late in the morning.

Steady afternoon shifts also have significant advantages. Often the worker arrives home at an hour when he or she can enjoy uninterrupted "quiet time" with a spouse or companion. The worker often has enough free time during the day to take care of outside business. The worker is able to sleep during normal hours and to sleep late in the morning. This shift also has major disadvantages for workers with families and significant relationships. A worker in such a shift will not have much time to see young children because they are asleep when the worker comes home, and at school when he or she wakes up in the morning. The major disadvantage of this shift for most people is that it interferes greatly with social activities.

While the steady night shift is often thought to produce the most detrimental effects, there are benefits available for the worker who can make the adjustment. This shift may actually interfere less with family and social obligations than the afternoon shift. Like the afternoon shift, working at night allows the worker useful free time and the opportunity to tend to outside business. Night workers seldom work fewer hours, but may get paid more and often do not have to

work as hard. The major disadvantage is having to sleep during the day. A night worker's demand for sleep in the daytime, when children are active and the home is noisy, can be a source of friction within a family. However, most permanent night workers, once conditioned, apparently do not experience major adjustment problems.

The rotating shift is the most commonly used form of shift work for a number of reasons. Without occasional relief, permanent afternoon workers would express dissatisfaction from being permanently cut off from family and social activities on work days. Shift rotation is used to spread the inconvenience around. The pay differentials commonly offered to night workers are not usually enough to compensate for the human costs of working nights as a permanent schedule. One solution is to adjust the timing of shift to minimize worker effects. The length of time between cycles of rotation could be lengthened to allow for adjustment of circadian rhythms. Similarly, increasing the frequency of vacations could allow for more free time between work periods. Minimizing the number of hours worked on a night shift could be beneficial, and the deficit could be made up on the other shifts. One group (cited in Mott) has suggested that the shifts should be adjusted to circadian rhythms relating to body temperature. The "dawn shift" would run from 4:00 a.m. to noon, the sunset shift from noon until 8:00 p.m., and the night shift from 8:00 p.m. to 4:00 a.m., thus allowing the night worker to get more sleep. This plan would meet with resistance from workers who do not like to rise in the dark early morning hours for the dawn shift, however.

Employees may sort themselves into the shift that suits them best. Employers that take into consideration personal preferences for shifts may allow those preferring the night shift to work it permanently. Certain personality types enjoy working at night and may lead a solitary life.

Younger workers, "evening people," extroverted personalities, and those with high-amplitude circadian rhythms appear to adjust more easily to shift work. As shift work becomes more the norm, the activities and services of the community could be changed to gear

activities to shift workers. In some communities dominated by a single employer that requires shift work, it is not unusual for services to adjust their hours.

Individual workers may find that the effects of shift work are mitigated by a routine for going home (preferably before sunrise if on a night shift) and going to bed at the same time each day in a darkened room or using an eyeshade, and sleeping at a comfortable but cool temperature. Avoiding alcohol and stimulants, including caffeine, helps to maintain a regular sleep cycle. Regular mealtimes also help, as does eating less and consuming mostly carbohydrates at the end of the shift in order to encourage sleep. Maintaining some consistency in daily routine, such as the approximate time of a major meal with one's family, can help mitigate the effects of shift work. Some workers find that little rituals, such as exercising, taking a walk, reading, playing with children, or cooking, helps to anchor their day when done consistently at certain times of the day.

Although some authors have attempted to relate shift work to morbidity, the evidence that shift work is a major risk factor for clinical disease is weak. Management of certain chronic illnesses, such as insulin-dependent diabetes and epilepsy, is certainly made more difficult by shift schedules and sleep deprivation, respectively. The most prominent effects of shift work are likely to be psychological and social. Shift work may not make most people clinically ill, but it can make their working lives very unpleasant.

### **Burnout**

Job burnout can be defined as a debilitating psychological condition characterized by a sense of disappointment, lack of fulfillment, and anxiety, brought about by unrelieved work stress. This may lead to a feeling of depleted energy reserves, increased dissatisfaction and pessimism, increased frequency of absence and diminished productivity when present, and lowered "resistance" to illness (whether through psychological response to minor illness, symptom magnification, or a true effect on the immune system).

There is only indirect evidence that stress on the job affects immune function and renders a person more likely to develop clinical illness, such as respiratory tract infections. However, most would agree that workers under great stress are less able to cope with illness. The fatigue, mental state, possible self-medication, and sleep disturbances may make them more susceptible to injury, substance abuse, and depression, as well as a wide range of psychosomatic disorders. It was once believed that the major risk factor in job burnout was the high degree, frequency, and intensity of intrinsic stressors in highly pressured individuals such as corporate executives, emergency response personnel, physicians, nurses, firefighters, police officers, social workers, and clergy. Occupations with few or no obvious sources of stress were considered low risk. In recent years, however, it has been observed that many jobs traditionally thought to have few or no sources of stress also experience job burnout among workers.

Job burnout may begin gradually and may vary from individual to individual. The worker may become bored or dispirited with the job, may become lackadaisical toward job responsibilities, may begin to procrastinate, and may fail to perform at a level compatible with past standards or personal potential. These patterns soon become noticeable to superiors. When the decrement in performance is brought to the worker's attention, the worker may become even more frustrated. As the worker becomes increasingly dissatisfied with the job, he or she may begin looking for any excuse to call in sick, and absence episodes may increase. Workers may develop psychosomatic illnesses or may experience exaggerated symptoms from real complaints that are rather minor, especially chronic pain. Sleep disorders are common and may show depressive features.

When the worker becomes frustrated with the job, there may also be acting-out behavior with those who are closest, such as spouses and children. Instead of confronting the problem at work, the worker may bring his or her feelings home, and complain to the partner or spouse and take out frustrations on the family. This behavior increases tension at home and may put family relationships in jeopardy.

The pace of work, within limits, is not the determining factor in job burnout. Workers who receive too few stimuli often feel that their job is monotonous and that they are not truly accomplishing anything. Most investigators now believe that a certain amount of controlled stress in a job is not unhealthy. Stimulation allows a person the opportunity to feel a sense of accomplishment and satisfaction and to feel better about themselves and their job. Feelings of depression are often more common among workers doing jobs of low complexity, and less common among workers in highly complex jobs. One useful study showed that among workers experiencing a sharply reduced workload, assembly line workers reported only slight dissatisfaction when they had too little to do, administrators reported significantly more, and police officers none. This response was correlated with the intrinsic interest reported in their work: administrators see themselves as challenged by problems and as indispensable to their employers, while assembly line workers saw themselves as replaceable human parts performing a routine function. Police, of course, deal with trouble, and for them satisfaction may be the absence of problems.

One associated factor in job burnout is what has been termed the "honeymoon effect," a sense of lack of fulfillment from the job that usually, but not always, affects those who are just beginning a new career after a prolonged period of preparation or education. The newly hired workers may begin very enthusiastically in the job, but soon may feel disappointed because they are relegated unchallenging duties day-in and day-out, or because their ideas are disregarded by their employers. They begin to feel that the time spent in preparation for their career was wasted. They lose their enthusiasm and soon feel that they are in a rut.

Deskilling is another factor in burnout. With increasing use of computers, mechanization of production, and unitizing and deskilling of work processes, many employers have created standardized, interchangeable jobs. These standardized jobs give the worker little opportunity for personal input, and lack sufficient variety for the worker to increase the level of skill in a particular workplace. This

leads to frustration and boredom in the absence of new challenges or accomplishments. When workers have been doing the same thing for many years and see nothing in their future except more of the same, they may feel trapped and unable to increase their status. They may adopt a “don’t care” attitude in defense, which often leads to job burnout.

Another factor involved with job burnout is working hours. Many jobs, especially those involving professional people or that have peak load of demand (such as waiters or bus drivers), do not accommodate a structured 40-hour work week. This leads to increasing social alienation and many of the same problems as one encounters with shift work.

Personality characteristics appear to play some role in setting the stage for burnout. Individual life situations are also important factors in determining risk. Individuals who are working below their educational status or skill level appear to be at increased risk of job burnout. They usually are of high intelligence, are not easily satisfied, and are demanding of themselves and others. Often they are stuck in jobs below their appropriate level because of personality traits. They may have a high image of themselves but have difficulty communicating how they feel to others, especially those in authority. They are very ambitious and wish to make a good impression, but may have difficulty expressing their ideas and suggestions to their employers, and soon feel frustrated. Conversely, they may have a low self-image and may feel there is nothing better that they can do, and that they are trapped in the job. These more susceptible individuals seem to be most vulnerable to job burnout either when starting a new job or after being in a single position for a long time.

One such group is popularly called “workaholics,” a colloquial term that implies a psychological dependency on work to alleviate anxiety or to avoid social problems. This is a manifestation of the familiar “neurotic paradox” in which a person engages in dysfunctional behavior that temporarily alleviates anxiety, but actually creates more anxiety in the long run. Workaholic behavior patterns may be better understood as work-dependent addictive behavior. Workaholics are typically individuals who have no other social fulfillment outside

their work. They become engrossed in their jobs as a means of retaining self esteem. Even play and socializing is tied to work and the familiar company of co-workers. Many ambitious people, in an effort to get ahead, put in considerable time beyond the usual hours, and may become totally engrossed in the job. When this becomes their only meaningful activity in life, the slightest setback may induce depression and feelings of inadequacy. Burnout comes not from fatigue but from depression, after the inevitable realization that he or she is unfulfilled, or after a career setback. The problem is compounded when these work habits interfere with family life, meaningful relationships outside work, and social roles in the community. Workaholic behavior tends to be very hard on families, not only because of the worker's absence, but also because the behavior is associated with emotional detachment and lack of engagement even when the person is at home. Recovery involves seeking other sources of satisfaction in life, building a social life that does not depend on co-workers, and behavioral change. It usually entails substituting more healthy behaviors for dysfunctional adherence to work.

However, workaholic behavior does not necessarily mean pathology. The behavior pattern must be considered in context. If the cost is reasonable in terms of family and social life, an intense working style is not in itself indicative that a person is a "workaholic." For some highly motivated people, intense work is a price paid now for gain later. Creative work can be a joy that one does not want to give up. Socializing with peers or people from the same work community is not in itself abnormal and is a characteristic of many careers, such as in academic life. As with most behaviors that border on addiction, the test is whether or not it interferes with life, and whether it can be changed without undue anxiety.

Another form of extreme work-dependent behavior is the Japanese phenomenon of "karoshi." The world of Japanese business has been intense, peer-enforced, and compulsorily social, with heavy peer pressure on young and mid-level executives to conform. It is not rare for mid-level managers, called "salarymen" because they draw a salary and are universally male, to work 70 hours per week, to engage in

after-work drinking and entertainment, and to spend long hours commuting (because affordable housing is usually distant from downtown business centers in Japan) and, consequently, very little time at home. In Japan, sudden death from heart disease is called “karoshi,” or death by overwork, and appears to be associated most closely with hypertension. However, karoshi victims also smoke more and have lower HDL cholesterol than age-matched co-workers. There is now a broad discussion and re-examination of this lifestyle in Japan, and growing recognition that it is unhealthy, unsustainable, and inimical to creativity.

Employee assistance programs (discussed in Chapter 19) may provide the support and interventions needed to resolve the problems of extreme work behavior. Discussing the problem with a counselor is often useful to clarify matters. Talking over problems at work may give new insights on how to handle them, as well as a renewed outlook on the job situation.

If these relatively modest steps do not help with the problem, further steps may be needed. A discussion or, if necessary, a confrontation with the supervisor may be unavoidable. If so, a counselor can help mediate and make the confrontation less personal or stressful. Rotating jobs or asking for a leave of absence may help. After taking time off, a worker may return with a more positive outlook toward the job. The worker may need a complete change of position or a new job in a new workplace. Recognition and early intervention are important in the reduction of job burnout. However, as with any other occupational disorder, the key to managing job burnout is prevention. In situations that allow it, rotating workers from time to time through different jobs will prevent monotony from setting in and makes for a more broadly skilled employee who will ultimately be more valuable to the employer. (This is often precluded by contract or union rules or is not otherwise feasible.) Seminars and workshops on stress awareness can be developed as part of a health prevention program. Promoting workers to higher-level positions from within, instead of filling higher positions with people outside the company, instills optimism and the feeling that a job well done will be rewarded.

Personal communications with employees is difficult for some managers, but is important and effective in improving worker morale. Acknowledging, in person and in memos, appreciation of the work that the worker has done makes the worker more satisfied with what has been accomplished and more eager to continue.

Employee input and suggestions, offered in good faith, should be received seriously and acted on if possible. Allowing the worker a chance to participate in the employer's growth by contributing ideas and suggestions makes the worker part of the team. This instills loyalty and makes for a stronger employee-employer relationship.

### **Stress Reduction**

Stress reduction for the individual may be achieved by environmental intervention to reduce stressors in the workplace, by stress management training, and by building resilience. Environmental intervention is generally under the initiative and control of management and applies across the workplace. Interventions for reducing stress in the workplace environment may be introduced at the group level, through changing work organization, or at the individual level, by means designed to help a particular employee. Stress management training, in the traditional sense, seeks to equip the individual with tools to manage his or her response to stress. Resilience training seeks to build the capacity within the worker to withstand and cope with reasonable levels of stress. A major goal of resilience training is to assist the worker in forming stronger social support systems.

Personal stress management programs are often included in health promotion programs and employee assistance programs (see Chapter 19). Table 13.3 outlines several general approaches to stress management in the workplace. These are general approaches and should be adapted as needed for a particular situation in the workplace. Different people have different stressors in their lives and different responses to the stresses they create, as well as very different lifestyles, social connections, and needs. The best solutions are therefore those that are appropriate for the individual.

**Table 13.3.** Selected Approaches to Managing Stress in Affected Individuals

*Taking Action*

- Taking control: setting goals, taking concrete steps to solve problems, sending out resumes for other jobs, etc.
- Identify and control particular stressor
- Awareness
- Modifications in work environment
- Speak to supervisor
- Union grievance
- Change jobs (if no alternative)

*Work Organization*

- Time out
- Holiday
- Days off
- Unstructured break during day
- Recreation or enjoyable activities on own time
- Schedule different work patterns
- Reassignment
- Telecommuting

*Behavioral Modification*

- Exercise (especially valuable for depressed mood)
- Relaxation techniques
- Autogenics (autosuggestion)
- Self-hypnosis or hypnosis
- Meditation
- Biofeedback
- Yoga
- Cognitive coping interventions
- Positive attitude: visualization techniques, self-talk, rehearsal
- “Minimizing” approach to coping style (reduce perceived seriousness)
- Decreased sense of vulnerability
- Improved coping skills
- Treatment of phobic behavior, when applicable

*Physical Maneuvers*

- Stretching

*(Continued)*

**Table 13.3.** (Continued)

---

Breath control
Progressive relaxation
<i>Social Support and Skills</i>
Social support and self-help networks
Conflict resolution training
Assertiveness training
<i>Clinical Intervention</i>
Individual psychotherapy
Control of substance use
Short-term use of tranquilizers and hypnotics (only to break pathological behavioral cycles, not for maintenance)
Control stimulant intake: coffee, tea, cola drinks
Control depressant intake: alcoholic beverages, prescription drugs
Identification and treatment of sleep disorders, GERD, other relevant conditions
<i>Lifestyle Interventions</i>
Diversion or awarding avocation or sport
Improved fitness, nutrition, and compliance with regular exercise regimen
Avoid hassles of daily life (to extent possible)
Regular bedtime

---

Work is a social environment and provides a worker with many rewards beyond the wage or salary. Most workers enjoy their work and deal with it on a daily basis with motivation and interest. One of the most rewarding aspects of work is a sense of belonging to a community with a shared purpose. The worker who does not enjoy a feeling of contribution may develop resentment that his or her contribution is not valued or a sense of diminished self-esteem, that he or she has little to contribute.

Maintenance of a positive worker attitude is critical to effective teamwork. Management can induce stress by actions that affect workers' attitudes toward the employer and the job. They can be pulled out of a "team." (Management may not perceive a particular group of workers as a team, but the workers themselves may.) They may be given repetitive work, denied due recognition, or compelled to meet what they deem to be extremely difficult and arbitrary

production demands. One of the most potent stressors in a work organization is the feeling that one is ignored or given no respect, and has no voice in the workplace or the organization as a whole.

Good managers are sensitive to the levels of stress in the workplace, to morale, and to the informal organization of the workplace. They empower workers by giving them a sense of involvement in the work and by listening to their input as decisions are made. Information flow in both directions is critical to maintaining responsive relationships between workers and their superiors. Communication must be fluid between the worker and those not only above, but also on the same level and below the worker. Not being able to talk with someone in the workplace about problems, especially to one's supervisor, is itself stressful.

Other managers are insensitive to the dynamics of the workforce and may not allow or encourage feedback from staff, which may also do harm to the employer by keeping out useful suggestions for production or problem solving. Managers with such attitudes may be unable to tell when employee morale is deteriorating until it is reflected in reduced performance and is harder to turn around.

Behavioral approaches to the management of psychological and physical responses to stress have been developed in the past decade. Relaxation techniques, assertiveness training, guided imagery, meditation, biofeedback, and exercise are useful for prevention as well as treatment. Table 13.4 summarizes some of the stress management techniques that can be used in individual cases.

Superiors can be made aware of the signs of stress, which may include changes in attitude, changes in work quality, absence from work, or reports of domestic troubles. Management, when requesting an increase in production with no increase in staffing or resources, should be aware of the potential for increased stress, and ideally should seek the means to mitigate it and to provide adequate resources. If an individual worker is identified as abusing alcohol or drugs or is affected by a mental health or dependency problem, he or she can be referred for help in an employee assistance program (see Chapter 19). If there is a physical health problem due to stress, the

**Table 13.4.** Selected Approaches to Reducing Stress in the Workplace

- 
- A “stress audit” to identify perceived sources of stress among workers and managers
  - Stress management training sessions for workers and supervisors
  - Supervisor training to control sources of stress
  - Employee assistance programs (see Chapter 19)
  - Health promotion programs (see Chapter 19)
  - Management restructuring and policies designed to reduce ambiguity, insecurity, and confusion over goals
  - Restructuring of the workplace for improved efficiency, worker convenience, and comfort
  - Restructure work to permit telecommuting or other flexible work practices (including means of monitoring productivity and work product)
  - Encourage participation of workers in planning their tasks and workplace design
  - Change work organization to provide greater variety, skills training, and social interaction
  - Share information on group goals and performance with workers
  - Examine the organization of work and consider where it can be liberalized or made more flexible
  - Actively seek opportunities for flex time
  - Allow workers to have input into decisions that affect them
  - Delineate job descriptions, goals, and career paths as explicitly and unambiguously as possible\*
  - Match the job to the worker’s demonstrated skills and capacity, with a “stretch” that is realistic and within the workers’ capabilities
  - Do not manage stress by releasing a worker from a task or requiring reduced hours or responsibilities\*\*
  - Examine the environment to determine if there are physical or ergonomic stressors that can be fixed
  - Identify sources of conflict within the workforce
  - Vigorously and visibly enforce policies against bullying and harassment\*\*\*
  - Keep meetings short and focused
  - Give workers enough time to do their work
- 

\*Workers, like all people, respond better when they know what is expected of them and how they can succeed. Ambiguity is profoundly stressful.

\*\*This sends a message that the employee is dispensable. Instead, whenever possible find a way for the worker to keep up, such as assigning an assistant, obtaining new software, or redefining the task.

\*\*\*Do not assume that every interpersonal problem is a personality conflict. Look for serious underlying issues and evidence of harassment and bullying before concluding that people just cannot get along.

worker should be evaluated by a physician who is an expert in that particular health field before the assumption is made that the problem is due entirely to stress. The occupational implications of any impairment should be evaluated by applying the fitness-to-work process (see Chapter 18).

Methods of job enrichment may also be used to increase job satisfaction. Job enrichment is the redesign of work to include tasks and activities that promote the psychological involvement of the worker in the work itself. This may include rearranging tasks and processes, adding new ones, increasing feedback on results and performance, increasing work variety, and increasing the level of contact with other workers.

A special type of occupational stress is associated with participation in or observation of a disaster such as an industrial accident or near-miss. This form of psychological trauma and the posttraumatic stress syndrome associated with it are discussed in Chapter 19.

## RESOURCES

- Cooper CL. *Handbook of Stress Medicine and Health*. 2nd ed. Boca Raton, FL: CRC Press; 2004.
- Sutherland VJ, Cooper CL. *Strategic Stress Management: An Organizational Approach*. London: Macmillan Business; 2000.
- Williams S, Cooper L. *Managing Workplace Stress*. New York: John Wiley & Sons; 2002.

## NOTEWORTHY READINGS

- Baum A, Garofalo J, Yali A. Socioeconomic status and chronic stress: does stress account for SES effects on health? *Ann NY Acad Sci*. 1999;896:131-144.
- Bosma H, Peter R, Siegrist J, Marmot M. Two alternative job stress models and the risk of coronary heart disease. *Am J Pub Health*. 1998;88(1):68-74.
- Cooper C, Dewe P. *Stress: A Brief History*. London: Blackwell Publishing; 2004.
- Cooper C, Robertson I. *Well-being in Organizations*. Hoboken, NJ: Wiley & Sons; 2001.
- Frone MR, Russell M, Barnes GM. Work-family conflict, gender, and health-related outcomes: a study of employed parents in two community samples. *J Occup Health Psychol*. 1996;1(1):57-69.

- Hubbard JR, Workman EA. *Handbook of Stress Medicine: An Organ System Approach*. Boca Raton, FL: CRC Press; 1998.
- Humphrey JH. *Women and Stress*. New York: Nova Science Publishers; 2004.
- International Agency for Research on Cancer. *Shiftwork, Painting, and Firefighting*. IARC Monographs. Lyon (France): IARC; 2008.
- Kahn JP, Langlieb AM. *Mental Health and Productivity in the Workplace: A Handbook for Organizations and Clinicians*. San Francisco: Jossey-Bass; 2003.
- Karasek RA, Baker D, Marxer F, Ahlbom A, Theorell T. Job decision latitude, job demands and cardiovascular disease: a prospective study of a random sample of the Swedish working population. *Am J Pub Health*. 1998;78:1335-1342.
- Karasek RA, Theorell T. *Healthy Work*. New York: Basic Books; 1990.
- Linden W. *Stress Management: From Basic Science to Better Practice*. Thousand Oaks, CA: Sage Publications; 2005.
- Lovallo WR. *Stress & Health*. Thousand Oaks, CA: Sage Publications; 2005.
- North FM, Syme L, Feeney A, Shipley M, Marmot M. Psychosocial work environment and sickness absence among British civil servants: The Whitehall II Study. *Am J Pub Health*. 1996;86:332-340.
- Peterson CL. *Work Stress: Studies of the Context, Content and Outcomes of Stress*. Amityville, NY: Baywood; 2003.
- Potter-Efron RT. *Handbook of Anger Management: Individual, Couple, Family and Group Approaches*. New York: Hawthorn Press; 2005.
- Rahe RH. Life-change measurement as a precipitator of illness. *Proc Roy Soc Med*. 1968;61:1124-1126. [Classic.]
- Seaward, B. *Managing Stress: Principles and Strategies for Health and Well-being*. Sudbury, MA: Jones and Bartlett Publishers; 2004.
- Selye H. *The Stress of Life*. New York: McGraw-Hill; 1956. [Classic]
- Smith D. Negative emotions and coronary heart disease: causally related or merely coexistent? A review. *Scan J Psychol*. 2001;42:57-69.
- Steptoe A, Evans O, Fieldman G. Perceptions of control over work: psychophysiological responses to self-paced and externally-paced tasks in an adult population sample. *Int J Psychophysiol*. 1997;25(3):211-220.

*This page intentionally left blank*

# 14 HAZARD CONTROL

A working knowledge of the available technical approaches helps physicians to deal with occupational health problems. The solution of an occupational health problem at reasonable cost usually requires knowledge and technical skill beyond the physician's training. Although the occupational physician is unlikely to become involved often in major engineering and plant maintenance decisions, an occupational and environmental medicine (OEM) practitioner needs to have a working knowledge of the principles of hazard control in order to be effective as a member of the health team. The physician also needs to share a common vocabulary with other occupational health professionals, particularly industrial hygienists and safety engineers. The basic approaches are summarized in Table 14.1. It is also helpful to know something about the other members of the occupational health and safety team, and their professional world.

Hazards in the workplace must first be identified, and then evaluated, before a strategy can be developed to control them. All three steps require specialized expertise. This three-part function is the responsibility of two specially trained occupational health professionals, occupational hygienists and safety engineers.

**Table 14.1.** Technical Approaches to Solving Occupational Health Problems

	Effectiveness	Cost	Reliability
Engineering controls	High	High	High
Containment and isolation	High	High	High
Personal protection	High	Moderate	Variable
Housekeeping	High	Low	Variable
Administrative controls	Low	Low	Variable
Behavioral controls	Low	Low	Low

Industrial (or in the UK, Canada, and other Commonwealth countries “occupational”) hygienists deal mostly with problems of exposure, ventilation, and plant processes. The emphasis in industrial hygiene is on reducing the exposure sustained by workers through controlling hazards and specifying effective protective measures. Certified industrial hygienists (CIHs) have backgrounds in chemistry or engineering and are trained in the technology of hazard control. Their expertise overlaps that of some physicians in toxicology, but their preparation in engineering controls makes them the experts in managing the hazard. It is almost always wiser for an OEM physician who may be advising a company to recommend a consultation with a hygienist than to make suggestions on control that are likely to be beyond his or her expertise.

Industrial hygienists in North America are experiencing some of the same changes in work organization that have been experienced by physicians. Large companies usually still have industrial hygienists on staff, but there has been a strong trend toward outsourcing hygiene functions in recent years. Some large companies, and most middle-sized companies in industries of low risk, use contract hygienists for all of their needs. Consultant hygienists with specialized expertise, for example in noise, are often retained on a short-term basis for particular jobs by medium-sized companies. These functions are becoming integrated into larger administrative functions that handle liability and loss control. Small companies are frequently at a disadvantage in not having specialized expertise

available. In recent years, hygienists have also diversified into environmental engineering, to become “bioenvironmental engineers.” However, that phrase usually implies management of the indoor environment in buildings together with industrial hygiene services. Industrial hygiene technicians, or “industrial hygiene assistants,” are trained to do a limited number of common, relatively simple procedures. They are both extending the reach of professional hygienists and competing with them for the low-end jobs.

Safety engineers are more common across the board in medium and smaller-sized industries, and deal mostly with physical hazards, fire control, and safe work practices. Increasingly, they are taking on some of the duties previously performed by hygienists. Safety engineering is less formalized as a profession than industrial hygiene, and salaries are lower. It is much more common for employers to have safety engineers on staff than to use consultants, except for special problems.

The identification and evaluation of hazardous exposures in a particular industry requires experience in that sector as well as special monitoring instruments that test for specific chemical and physical hazards likely to be found in the workplace. Some problems are simple to solve, but many require extensive and costly changes in the plant structure and the manufacturing or production process. Selection and design of the best solution to the problem can be a highly technical challenge to the engineer and plant manager. A teamwork approach is required for this to turn out well.

### **Controlling Occupational Hazards**

The basic approaches to controlling occupational hazards fall into five general categories: engineering controls, containment and isolation, personal protection, housekeeping, administrative controls, and behavioral controls (see Table 14.1). There are variations to these approaches and numerous “tricks of the trade,” but an understanding of these basic approaches will help the physician understand what solutions may be proposed for a given occupational health problem. Table 14.2 provides tools and approaches that the hygienist can use within a broader hazard-control strategy.

**Table 14.2.** Specific Measures for Controlling Exposure to Hazards

	Category*
Elimination	E
Substitution	E
Process change	E
Product change	E
Isolation	CI
Enclosure	CI
Housekeeping	H
Dust suppression	H
Maintenance	E, H
Sanitation	E, H
Waste disposal	H
Work practices	B
Education	B
Labeling and warning	B
Personal protection	P
Environmental monitoring and response	A, E
Personnel tracking and assignment	A
Program management	A
Hazard inventory and tracking	A
Occupational health audit	A

\* A = administrative, B = behavioral, CI = containment/isolation, E = engineering controls, H = housekeeping, P = personnel protection

Table adapted from Dr. C. Billings, Massachusetts Institute of Technology.

The optimal solution to a particular problem balances the technical requirements with the available resources and may involve a blend of approaches.

### ***Substitution and Elimination***

Hazardous chemicals, machinery, and processes may not be necessary to manufacture a product or deliver a service. It may be possible to

re-engineer the process to remove a hazard or to substitute a less hazardous element, such as a less toxic solvent. An entire technology of “green chemistry” is under development to substitute nonhazardous and environmentally sustainable alternatives for chemicals currently used in industry.

### ***Engineering Controls***

Engineering controls can be very expensive, especially when they are retrofitted to older equipment. Engineering controls are usually more effective and economical when built into the original design of the plant. Improper maintenance and operator neglect may result in reduced effectiveness of even the best controls, however. An example of an engineering control is the design of heavy equipment to minimize noise and vibration, and the mounting of machinery on sound-absorbing pads.

### ***Containment and Isolation***

Containment may involve the construction of a shell around a machine or placing a barrier between the source and the work area. Isolation is the removal of a particularly hazardous process to a location away from workers. Both are highly effective ways of dealing with many hazards. Chemical processes are usually tightly contained in an enclosed tank or reaction chamber, for example. Sound-absorbing baffles or enclosures are commonly used to reduce noise exposure. As with engineering controls, however, the effectiveness of containment and isolation often depends on careful maintenance, and is subject to carelessness. Care must also be taken that an isolated facility does not become an environmental hazard to the surrounding community.

### ***Personal Protection***

Personal protection is generally considered to be the last option for reducing exposure, to be used when engineering controls are impractical. Personal protection places a barrier between the worker

and the worker's immediate environment; it does not eliminate the hazard. If personal protection fails, there is no backup, and uncontrolled exposure may result. The effectiveness of control devices depends on the fit between a personal protective device (such as earplugs or a respirator) and the worker using it, as well as on the education and compliance of the worker, and the commitment of the employer in ensuring proper use of the control device. The selection of the proper protective apparatus for the job should be done by the expert. However, whenever possible, a selection of alternative devices from which workers may choose should be provided, because this will enhance compliance as well as the effectiveness of personal protection. Some hazards can be dealt with only by personal protection; work outdoors is one such example, but personal protection is generally not the best approach if engineering controls are possible and feasible. Personal protection is affected by too many technical and behavioral variables. The user must be trained in its use, must be motivated to use the protection correctly, and must maintain the equipment.

### ***Administrative Controls***

Administrative controls are used primarily when cumulative exposure should be kept to a permissible level, as with low-level ionizing radiation. Workers may be rotated among stations on a regular schedule to reduce the possibility of any one individual receiving excessive exposure. Administrative controls are useful as an added safety precaution, but rotation is not a legitimate substitute for controlling the exposure itself. In emergency situations, usually involving radiation, administrative controls may be necessary as a temporary measure until the problem is brought under control. However, as a primary strategy for an exposure, such as noise, that can be controlled by other means, this solution is not ethical.

### ***Behavioral Controls***

Behavioral controls involve the education and motivation of workers to minimize exposure and to practice safe work habits. They may also involve the posting of warning signs or signals. This is

usually considered the least satisfactory approach because it depends on the individual worker's alertness, motivation, and comprehension and may be compromised easily by minor distractions, boredom, fatigue, passive-aggressive behavior, ill health, poor eyesight or hearing, misunderstanding, language barriers, and lack of personal motivation. Training for safe performance of the workers' duties and educational programs that broadly inform the worker about hazards and raises awareness are always helpful. However, protection from a workplace hazard should not depend exclusively on behavioral controls, because they are fallible. They depend too much on the worker's training, comprehension, motivation, attitude, and attention.

### ***Housekeeping***

Usually not separately identified, housekeeping is an effective approach to controlling many potentially hazardous exposures. Many hazards accumulate as residues of the work process, such as lead particles contaminating dust in a plant, fiberglass dust in a warehouse, and chemicals staining or contaminating work surfaces. Others accumulate as wastes in waste storage areas, as unused chemical inventory, or as discards in trash. In the absence of effective housecleaning, some of these hazards may pose a threat to workers in operating areas. (In one manufacturing plant, for example, lead-containing dust regularly rained down from the inaccessible rafters on workers below when the drop hammer vibrated the building.) Even when housekeeping is adequate, maintenance and cleaning personnel may be placed at some risk. Generally, housekeeping represents an effective but often overlooked general measure that can significantly improve both working conditions and the potential for exposure to toxic agents. The activities of cleaning and maintenance also often bring problems to attention.

## **PERSONAL PROTECTION**

Whenever possible, hazard control should be accomplished by engineering controls, including isolation and containment; if this is not possible, airborne exposure should be controlled by ensuring

adequate ventilation. In many situations of emergency response, upset conditions, occasional access, or maintenance operation, it may not be feasible or possible to control the hazard at its source. Personal protective equipment (PPE) may then be used.

The OEM physician should be very familiar with PPE, including function, indications, effectiveness, compliance patterns in the specific workplace, medical limitations, and issues associated with their use. Respiratory devices must often be tested for fit to a particular worker. Knowledge of whether or not a worker used PPE is an important part of the occupational history.

Table 14.3 lists the types of personal protection in common use in industry. The physician is most often involved in issues involving respiratory protection, hearing protection, and hand protection.

**Table 14.3.** Personal Protective Equipment in Common Use

Head protection	Hardhat, helmet, cloth cap, hairnet, hood
Eye protection	Goggles (UV, IR, clear), safety glasses, face shield
Respiratory protection	Dust mask, organic vapor mask, full-face canister respirator, emergency compressed-air bottle supply respirator, respirator spectacles
Hearing protection	Earplugs, earmuffs, acoustic helmet
Hand protection	Barrier cream, gloves, mitts, gauntlets
Foot protection	Safety boots or shoes, instep protectors, cleats, insulated boots, ground shoe covers, rubbers
Body protection	Impervious suit, waterproof garment, apron, heatproof suit, flameproof/-resistant garment, jacket, body armor, knee or elbow protection, wet suit

## Respiratory Protection

PPE for respiratory protection is based on “respirators” (not to be confused with ventilators, as in an intensive care unit). Respirators are of two types, air supply and air purifying. The different categories of respirator are listed in Table 14.4; they differ greatly in efficacy, degree of protection, and suitability for particular situations. Workers who may use respiratory protection equipment, as well as their supervisors and managers, must be aware of the limitations associated with each type of equipment and should be trained not to rely on inadequate devices. Simple cloth masks, for example, are unsuitable for use in any potentially toxic exposure. If a worker must enter an atmosphere in which the concentration of a potentially toxic exposure or the level of oxygen is unknown, the only acceptable

**Table 14.4.** Types of Respirators in Common Use

Major Use	
<i>Air Supply</i>	
Self-contained breathing apparatus (SCBA)	Emergencies, highly toxic atmospheres, escape, hazard unknown, firefighting
Air line with SCBA	Sustained work, escape possible
Air line	Sustained work, particulates, painting
<i>Air-Purifying</i>	
Positive-pressure respirators	Sustained work, low-level gases, hot and humid conditions
Negative-pressure respirators	
Full mask	Low-level gaseous, particulate exposures
Half-mask	Low-level gaseous, particulate exposures, not highly irritating to eyes
Cloth masks	Large particles

respiratory protection is a supplied-air system, either from an air line or self-contained breathing apparatus (SCBA).

### *Air Supply Devices*

The greatest degree of protection is given by air supply devices, which provide the wearer with a safe atmosphere from either a tank (SCBA) or an air hose (air line). Air supply devices, properly rated and tested, are used for atmospheres more than 50 times the PEL or STEL (see Chapter 8). (At levels more than a 1,000 times the STEL, a full-face piece or helmet and protective suit may be required.) All air supply devices should have full-face masks for complete protection. Air supply systems are absolutely required in any situation where there may be an oxygen-deficient atmosphere or in which carbon monoxide may be present at elevated concentrations. In addition to supplying oxygen, air supply devices displace the hazardous atmosphere in the worker's breathing zone, so that it is not inadvertently inhaled. As long as the face mask is tightly sealed, there is no risk of intrusion or breakthrough. Obviously, even air supply systems cannot be used in atmospheres that are highly irritating or corrosive unless the worker's body is completely protected.

Supplied-air respirators using air lines are more useful for prolonged work in situations where emergency rescue is not an issue, in confined spaces, and in atmospheres that are not immediately life threatening. (Some air line systems provide a small accessory SCBA tank for safe escape if the air line is compromised.) The worker wears a hood and often a full-body suit (for example, if sandblasting or spray painting) and air is supplied by a hose from a compressor. Supplied-air respirators may provide constant flow or may have a pressure-demand regulator. Constant flow has the advantage of cooling the worker in hot environments. Each part of the system should be well maintained, but it is particularly important that the hose and compressor be clean and uncontaminated.

SCBAs are generally preferred for work that requires mobility, in emergencies such as fires, or when the degree of hazard is unknown.

They are standard equipment for firefighting, mine rescue, and escape. SCBA tanks have limited capacity and impose considerable ergonomic demands on the wearer because of their weight. Recent advances in the technology of SCBA devices have reduced the weight and bulkiness of the apparatus, but even tanks made of light-weight composite materials may weigh 15 kg or more when filled. The self-contained underwater breathing apparatus (SCUBA) is a different proposition altogether. On land, there is no buoyancy to help the wearer carry the device.

In both types of air supply devices, care must be taken to ensure that the air supply itself is uncontaminated and that the supply lines are clean. Occasionally, accidents or problems occur because lines or tank are contaminated by oil from a compressor seal leak, or are supplied from a contaminated source, such as a garage in which vehicles are discharging exhaust, a workshop in which solvent fumes are strong, or a pump close to the emission source of carbon monoxide. Once compressed in a tank, and with no possibility for the wearer to take a breath of clean air after entering into the hazardous atmosphere, contamination of the air in the tank presents a serious hazard.

### ***Air-Purifying Devices***

Air-purifying devices rely on filters and chemical packs (typically in canisters of metal or plastic) to remove the hazardous chemical from the air delivered to the breathing zone of the worker. They are effective for relatively low levels of airborne hazards, at 10 times the PEL or STEL. These devices cannot be used where there is a risk of oxygen deficiency, as in a confined space.

There are two types of air-purifying devices: positive- and negative-pressure.

Positive-pressure devices are those in which a small pump draws air in through the purifying system and delivers it to the breathing zone of the worker at a positive pressure relative to the atmosphere and at a rate greater than the worker's ventilatory capacity. With these devices, the excess, unbreathed air flows out underneath the

mask and prevents the intrusion of contaminated air. These devices are comfortable, even in conditions of high temperature and humidity, and do not require a tight seal to the face. They are expensive, however, and not widely available. They are not suitable for highly toxic atmospheres, including those that contain high concentrations of carbon monoxide or cyanide; they are also unsafe to use with some other hazards, and are unsuitable for hazardous atmospheres in which the warning properties of the gas are weak, or for emergency use.

Negative-pressure respirators are face masks with a tight seal in which the wearer breathes through canisters, usually mounted to either side of the face mask, that filter dusts or purify the air through chemical changes specific to the chemical hazard. They are so named because the negative pressure of the wearer's inspiration draws air into the apparatus. Choosing the right type of canister for the specific hazard and ensuring a tight fit on the face are the keys to effective protection in using negative-pressure respirators. Their use is limited by the time to saturation of the canister material and the level of concentration that the canister is rated for effectiveness. If the canister fails or is overwhelmed in a hazardous atmosphere, the resulting exposure could be very dangerous and may catch the worker unawares because he or she assumed that the protection level would be adequate.

Negative-pressure respirators come in both full-face and half-face varieties. Full-face respirators provide a more reliable seal and protect the eyes, but they restrict vision and are cumbersome. Half-face, negative-pressure respirators are more comfortable unless an atmosphere is irritating to the eyes, and they provide unrestricted vision; however, the margin of safety is much lower.

Negative-pressure respirators come with detachable and replaceable canisters, color-coded for appropriate use. Canisters are designed for classes of chemical hazards, such as particles, hydrocarbon vapors, and reactive chemicals (such as ammonia and chlorine). They are not effective for highly toxic atmospheres, carbon monoxide or cyanide, and some other hazards, and they are unsuitable for emergency use.

The canisters have a finite useful life, after which they must be replaced. The mask itself should be kept clean and stored in a locker that is not dusty or contaminated. These devices should never be used in dangerous atmospheres with poor warning capacity—where there would be little warning of failure, as by a strong odor or subtoxic irritation. For example, an atmosphere with variable or relatively high levels of hydrogen sulfide, with its effect of olfactory paralysis limiting the warning that workers would have of imminent danger, would require SCBA. This is also true of positive-pressure air-purifying devices. They are, however, useful devices for personal protection in most routine situations.

### ***N95 and Other Limited Protection Devices***

N95 respirators are lightweight, disposable micropore respirators designed to protect the wearer from aerosols down to 0.3  $\mu\text{m}$ . They are very popular for nonspecific dust control. Surgical N95 respirators are extensively used in hospitals for protection against infectious hazards, such as splatter from body fluids, droplet nuclei carrying tuberculosis, or body fluid aerosols carrying HIV or other infectious agents. They are regulated by the FDA for this use. Concern for SARS, pandemic influenza, and bioterrorism has led to extensive stockpiling of N95 respirators by hospitals and public health agencies. N95 respirators are designed for a single use and should never be reused or washed. Their effectiveness also degrades when they are wet.

A powered air-purifying respirator (PAPR) is a battery-operated air-purifying device that blows air through a HEPA filter. PAPRs are worn with half-face masks. They do not require fit testing. PAPRs are used in hospitals and healthcare institutions when providers cannot wear N95 respirators because of poor fit, facial hair, or other limitations. PAPRs are not proper positive-pressure devices because a strong inspiration can overcome the positive air flow and entrain unfiltered air from around the mask. They do not provide the same level of protection as positive-pressure devices.

### ***Respirator Fit Testing***

For negative-pressure equipment, air lines, or SCBA devices to work properly, the mask must fit tightly over the face. Facial physiognomy varies greatly, and facial hair may prevent a firm seal. For this reason, it is usual in plants or industries where toxic exposures are possible to require that the workers be clean-shaven and undergo medical evaluation to ensure safe fit. Certain facial abnormalities, such as a cleft lip or a hypoplastic nose bridge, may preclude the use of a face mask. Some persons find that masks of one design or another press painfully on facial nerves.

Other factors are also important determinants of the suitability of a respirator. Negative-pressure respirators increase the work of breathing, because of inspiratory resistance. They may increase CO<sub>2</sub> retention slightly by adding to respiratory dead space. These factors change the pattern of breathing and may cause a sensation of dyspnea. During maximal exertion and stress, as in an emergency, the sensation of dyspnea may be disconcerting or frightening. Face masks also trap heat and may become uncomfortable in hot environments. Heat may result in a skin rash under the mask, especially under humid conditions. Vision may be reduced in a full mask, especially at the periphery, although the eyes are better protected. Contact lenses are never compatible with using a respirator, because of the irritation or toxic nature of the atmosphere in which they are used, as well as the risk of dislodging the lens inadvertently in a crisis by rubbing the eye. Positive-pressure respirators dry out contact lenses quickly because the air cannot be adequately humidified. Lenses can be placed in the visor of full-face masks, however.

Respiratory protection devices should be inspected periodically. The canisters and filters must be changed as often as required to ensure protection. Regular cleaning of the mask and hoses with soap and hot water should be the responsibility of one designated person, and should be performed after every use. Organic solvents and strong detergents should not be used because residues may persist inside the mask. Canisters and filters should always be kept dry.

Medical contraindications to use of a respirator probably include any chronic or acute respiratory disease that increases airway resistance or is associated with cough, moderate to serious cardiovascular disease (in which case the worker should not be in a position associated with the need for a respirator), claustrophobia or panic attacks, and facial anomalies precluding a tight seal. (In the past, a perforated tympanic membrane was considered a contradiction on the grounds that toxic gases could gain entry to the respiratory tract through the Eustachian canal, but experimental evidence has now shown that this is not a risk.)

Criteria for respirator fit testing follow the OSHA Fit Testing Procedures (29 CFR 1910.134). Any OEM physician who performs fit testing must read and become intimately familiar with the provisions of this standard. It provides for a standardized test that includes the following elements:

- Criteria for comfort
- Criteria for fit, with respect to positioning, placement of the chin strap, and prevention of slippage
- Room for eye protection over a half-face mask or within a full-face mask
- Stability with motion of the head up and down and side to side, grimacing, and bending over
- Comfort and ease of air exchange with normal and deep breathing
- Speech audibility and clarity
- Qualitative fit testing for gases using isoamyl acetate
- Qualitative fit testing for aerosols using chemical detectors: saccharine solution aerosol, irritating smoke (stannic chloride), or Bitrex®, an intensely bitter-tasting but safe chemical also used as a product additive to prevent ingestion and poisoning

Qualitative fit testing tests are usually performed in a sealed room or booth reserved for the purpose. While wearing the respirator, the

worker is exposed to an odorous test substance such as banana oil (isoamyl acetate) or Bitrex®. If he or she can smell or taste the test substance, the fit is unsatisfactory. This test can also be performed using a mildly irritating substance (such as stannous chloride) but this is often objectionable to the worker subjected to the test.

Advanced medical evaluation protocols for negative-pressure respirators and for evaluating tolerance in more complicated situations are still under development. Pulmonary function studies (especially the otherwise rarely used maximal volume of ventilation, for which tolerance rating schedules exist based on anticipated work exertion) provide something of a guide, as do cardiac stress tests for those who must function at maximal exertion. Stress testing has also been adapted by Phil Harber into a submaximal PPE tolerance test that can be performed on workers wearing the appropriate device at a medical office or at the worksite.

Quantitative fit testing is a method of objectively determining the degree of protection respiratory protective equipment provides. It is used to evaluate respirators and when it is necessary to document the level of protection when the risk is greater. Quantitative fit testing is prescribed in detail in the OSHA Respiratory Protection Standard, CFR 1910.134.

Quantitative fit testing is performed with a device (the main one in current use is the PortaCount®) that compares the concentration of aerosol outside the respirator with the air inside the mask to obtain a reduction factor, expressed as a ratio which is generically called a “protection factor.” There are actually three important “protection factors,” which should not be confused:

- Fit factor (FF), the ratio of aerosol outside to inside the respirator obtained from an actual fit test and specific to that one fit test. A FF of 100 would normally be considered acceptable for routine risks but higher levels of protection may be required.
- Assigned Protection Factor (APF), the level of protection, expressed by the same ratio, that a respirator can be expected to provide 95% of the times it is used, as determined by either NIOSH or the American National Standards Institute (ANSI).

The APF is a guide to safe use of a respirator because it can be used to determine whether the OSHA-mandated permissible exposure limit (PEL) will be exceeded by the person wearing the device. If the concentration outside the respirator divided by the APF exceeds the PEL, the device does not provide adequate protection. A typical half-face negative-pressure air-purifying respirator would normally have an APF of about 10 where a full-face negative-pressure air-purifying respirator might have an APF of 50.

Workplace Protection Factor (WPF), the level of protection, expressed by the same ratio, experienced under actual working conditions when the respirator is being used, not under ideal conditions. A WPF is normally less than the FF. The WPF can be measured in the field and is important in research on respirator effectiveness, especially to derive the APF and to compare products, but is not routinely used in practice.

The FF should normally be about 10 times the APF, in order to provide a margin of safety for conditions of actual use. (This is custom, not specified by regulation.) Quantitative fit testing is required when a FF of greater than 100 is required. Tight-fitting powered supplied-air respirators, both air-supply and air-purifying types, should be quantitatively fit-tested if they are going to be used in negative pressure mode, (with the pump or compressor turned off). Quantitative fit testing of N95 respirators is currently optional but recommended by the Centers for Disease Control for protection against airborne pathogens. Workers who request respiratory protection but are working in atmospheres that are in compliance with the PEL do not need to be fit tested but should be trained in respirator maintenance and use. Fit testing should be repeated every few years and when there are changes in the facial structure (for example, because of surgery or weight change).

Respirator fit testing should be combined with training on hazards, the limitations of respiratory protection, respiratory maintenance, cleaning, inspection, and disinfection (if multiple users). Respirator maintenance is not difficult but is easily forgotten or neglected.

SCBA tolerance evaluations are more rigid because of the extreme nature of the circumstances in which they are used: FEV<sub>1</sub>/FVC between 45 and 60 percent. FEV<sub>1</sub> must be at least 60 percent and FVC at least 50 percent predicted, or there must be a contraindicating factor. Otherwise, those between 50 and 60 years of age should be restricted to using respirators only in emergencies. In practice, "restricted use" means that they may be assigned to areas in which respirators are only needed occasionally or may be needed briefly in an emergency, but not in situations where use of a respirator is frequent or essential for protection of life. The majority of workers, being healthy, can be certified for unrestricted "full use." Those who do not meet the medical standards above should be declared medically "unfit" to use respirators and should not be assigned to any area where their use may be required.

Even in the absence of specific medical conditions interfering with the use of respirators, 10 percent of workers simply cannot tolerate them, for psychological reasons. Workers certified to use respirators should be reevaluated every three years, and more often as they age beyond 45 years if their work is strenuous.

There are simple tests that workers should perform every time they put on their PPE:

- *Positive pressure.* The exhalation port of the respirator is occluded with the hand and the wearer exhales. If the mask can retain a slight but noticeable positive pressure without breaking the seal, it is probably adequate.
- *Negative pressure.* The inlets of the canisters, filters, or hose are occluded with the hands and the wearer inhales and holds the breath for ten seconds. If the mask collapses slightly but the vacuum is maintained, the fit is probably adequate.

## Hearing Protection

The technical rationale for hearing protection is outlined in Chapter 9. Noise-induced hearing loss is discussed in Chapter 16. Plugs and muffs remain the most common types of hearing protection in use. Several types of hearing protection are in common use:

- “Earplugs” come in many varieties, including deformable foam, sponge, and spun fiber varieties. The disposable earplugs should not be reused or adjusted in the ear after insertion. Permanent earplugs are easily lost and must be kept clean. Plugs are contraindicated if the worker has ear problems, such as external otitis. Plugs of sufficient noise reduction rating (NRR) work well, with good noise attenuation characteristics. They should be inserted with clean hands in order not to introduce dirt into the ear. Commercial spun-fiber plugs are never made of cotton, which is ineffective as a sound-damping agent.
- Molded plastic earplugs may or may not be customized, and are usually permanent rather than disposable; they occlude the external auditory canal. They must be kept clean to avoid irritating the external auditory canal. Unless a worker has an unusual size or abnormal ear morphology, custom-fitted earplugs are unnecessary.
- “Earmuffs” are soft, usually plastic cushions that fit over the external ear, and are usually kept in place with a flexible band. If the muffs fit over the ears and seal well, they provide a good level of protection. The headband fits poorly under a hardhat, however, so earmuff hearing protection is often attached to a hardhat. It is harder to obtain a good seal and noise attenuation from these combined hardhats, however. Muffs must fit rather tightly—and comfortably—over the ear with a firm, airtight seal in the postauricular area and zygomatic arch. Some workers with prominent facial or neck features cannot attain a reliable seal with muffs. Under hot and humid conditions, continuous wear of muffs may become uncomfortable or may result in dermatitis.
- “Ear caps,” often called “canal caps,” are foam discs that fit into the ear but do not enter the canal. They occlude only the external meatus. They are a useful alternative in low-risk situations when a worker cannot tolerate irritation of the canal.
- “Noise-canceling earmuffs” and helmets use a rapid-response amplifier and tone generator that detects continuous incoming noise and then cancels sound waveforms by generating an

inverse waveform. One advantage of noise-canceling systems is that because they act on continuous noise, speech gets through clearly. However, impulse or intermittent noise is not screened out, so the device provides only partial protection.

The key to evaluating the efficacy of any hearing protection device is the noise reduction rating, or NRR, a figure that is derived from the sound attenuation coefficient, which represents sound volume reduction expressed in decibels (see Chapter 9). These laboratory-derived ratings are generally considerable overestimates (by as much as 100 percent) over actual noise reduction, in the frequency ranges and patterns found in workplace conditions. However, the NRR is a useful guide for comparing one type of hearing protection to another. It should be noted that rolled cotton, a common folk substitute for proper hearing protection, has an NRR so low that it provides effectively no protection.

Not everyone can tolerate the same type of hearing protection, nor is one type suitable for every job. Personal hearing protection should be only one aspect of a more comprehensive occupational hearing conservation program. Because there are so many individual differences among workers, and because personal preference plays such an important role in compliance with using hearing protection, it is strongly advised that at least two types or models of adequate hearing protection be available in noisy workplaces to allow workers to make the choice themselves.

Some workers have medical conditions that interfere with the use of certain types of hearing protection, including external otitis, malformations of the pinna, or impacted cerumen. In such situations, the underlying disorder should be evaluated and corrected, and the type of hearing protection least likely to aggravate the condition should be used. Some workers find that they are disturbed by tinnitus or changes in the quality of sound when they use hearing protection. Such complaints may result from the unmasking of an underlying tinnitus (the same reason tinnitus appears to be louder at night) or as a result of the “occlusion effect.” The occlusion effect occurs when the external auditory meatus is occluded.

Bone conduction is then perceived as louder than air conduction, and sounds at frequencies below 2 kHz sound louder. Adventitial sounds such as breathing and chewing sounds are exaggerated compared to environmental noise. The occlusion effect indicates that the hearing protection device is appropriately fitted and is therefore a useful sign. If it is too disturbing to workers, however, it can be substantially reduced by either inserting the ear plug more deeply into the auditory canal or choosing a muff with a larger inside volume.

The occlusion test is a useful indicator of a good fit. Another method is to cup one's hand over one's ears. If there is no markedly perceptible change in the intensity of perceived environmental noise, the fit is proper.

### **Skin Protection**

The principal routes of entry for chemicals encountered in the workplace are by inhalation and skin contact. Respiratory protection requires attention to a relatively small space, the breathing zone. Skin protection, on the other hand, requires protection of the entire surface area of the body, with particular attention to those parts most likely to come into contact with the hazard. Although light coveralls, jeans, T-shirts, and other typical work clothing may be sufficient for many jobs in which the only exposure is dust or grease, they provide little or no protection in workplaces in which the potential exists for exposure to toxic chemicals or dusts that present a toxic hazard.

The purpose of skin protection is not only to prevent injury to the skin itself, but also to prevent absorption through the skin of potentially toxic chemicals (see Chapter 2). The extent of absorption depends on the surface area of the skin exposed, the concentration of the chemical, the lipid solubility of the chemical or the vehicle in which it is dissolved, and the duration of contact. Absorption by the transdermal route is enhanced when the skin is inflamed, excoriated, or abraded, and when there are open wounds. Irritation caused by one chemical may promote systemic absorption of that chemical as

well as others. Lipid-soluble chemicals, such as solvents and many insecticides dissolved in petroleum-based vehicles, may penetrate the skin of most parts of the body very readily. Absorption is much faster and more efficient across mucous membranes as well as where the skin over the site of contact is thin and heavily vascularized: the face, scalp, back of the hand, perineum, and genitalia are particularly efficient sites of absorption. Thus, changing underclothes and maintaining personal hygiene are basic elements of personal protection against skin absorption.

Skin absorption of some toxic chemicals, such as parathion, is sometimes sufficient to cause serious systemic toxicity. In many other circumstances, exposure by skin contact and inhalation occur simultaneously, and absorption across skin may add to the total internal dose and contribute to the toxicity that primarily results from exposure by inhalation. This is a particular problem with pesticides and solvents.

### ***Handling Contaminated Clothing***

Protective clothing must be maintained and managed; it is not sufficient to provide a set of clothes and a locker and to forget about personal protection thereafter. Protective clothing must be kept scrupulously clean. If gloves, boots, or coveralls become contaminated on the inside surface, they may actually promote absorption. Protective clothing should be kept and washed at the workplace, not taken home or to a laundromat, so that chemical residues will not be carried home on the clothing, passively exposing family members and others. Passive exposure due to contaminated clothing at home has been a serious problem in the past, with cases resulting in significant exposure to asbestos and lead.

Contaminated clothing should be removed in a contaminated environment, not brought into an otherwise clean area. The clothing should be kept separate from clean clothes or, if it is necessary to take clothing out of the area, should be sealed in plastic bags.

Washing procedures may vary for individual chemicals. It has been suggested (by Betty Crown, of the University of Alberta) that the following is a safe washing procedure for pesticides, and

is therefore likely to be safe for other potentially toxic organic compounds:

- Pretreat heavily soiled areas with stain remover.
- Load the clothes into an automatic washing machine loosely, without overpacking.
- Never mix contaminated with non-contaminated clothing.
- Presoak the clothes.
- Use a full normal cycle on the hot water setting with a heavy-duty detergent.
- Wash clothes two or three times.
- If possible, dry on a clothesline to accelerate photolysis and the degradation of remaining residues, and to avoid contaminating the dryer.

### ***Hand Protection***

The skin of the hands can be protected by hand washing, barrier creams, and wearing gloves. However, each approach has its unique benefits and problems. Hand washing is an effective and important method of reducing contamination on the skin of the hands, reducing the potential for irritation and for absorption. It has the additional benefit of reducing the potential for hand-borne contamination of other parts of the body, of inadvertent ingestion of contaminants on the hands, and of contamination of the work environment. It is not unusual for workers to wash their hands several times a day with soap and water or, in a practice that is ill advised but has been almost ubiquitous, in solvent. Unfortunately, frequent hand washing also removes protective natural oils from the skin and may cause chapping and dermatitis, resulting in reduced natural protection from chemical exposure. Washing in solvent, of course, not only results in transdermal exposure to the solvent itself, but also elutes fats from the skin, and therefore may cause a serious dermatitis and greatly increased skin absorption of other chemicals. Alcohol-containing skin disinfectants also promote de-fatting and dermatitis.

Hand washing should be performed only using a mild soap and water, certainly at the beginning and end of the day and before eating, to prevent contamination of food. More frequent hand washing is required in certain occupations, such as nursing, and these occupations typically have very high rates of dermatitis. Hand creams should be provided at each hand washing station or sink in such situations, for immediate skin care.

Barrier creams are preparations applied to the hands that dry into a transparent, flexible, temporary, relatively impermeable barrier to water and many solvents. Some barrier films are said to turn into a membrane, with micropores that allow perspiration to evaporate without admitting water from outside; others are physical barriers consisting of silicon grease. Barrier creams provide useful but only partial protection in situations where gloves are impractical, such as in fine work using small parts. They cannot be relied upon to protect the hands against strong acids, alkalis, or prolonged exposure to solvents or reactive chemicals dissolved in water. They may be worn on the hands under gloves, as additional protection against chemicals that may penetrate glove material, to prevent irritation from moist glove linings, and to reduce the frequency of hand washing that would otherwise be required. Barrier creams can be washed off after work with soap and water. In choosing a barrier cream, one should look for products that are hypoallergenic, not greasy, and free of fragrances or other additives that may cause their own problems of sensitization, irritation, and contamination of food.

Gloves are the most effective form of hand protection, when properly selected and used. They have certain intrinsic disadvantages, however, that sometimes lead workers to avoid them. Gloves are often hot and impede evaporation of perspiration, causing the skin of the hands to become uncomfortable and damp over time. They are easily contaminated on the inside when workers put their gloves on after making skin contact with chemicals; when this happens, gloves may become like occlusive dressings, sources of irritation and absorption rather than protection. Latex, which has been used indiscriminately in many applications where it is not required, carries a high risk of sensitization and dermatitis. Hand size varies greatly, and gloves must be fitted to the person using them; a range of sizes should

therefore be provided to suite potential users. Depending on the thickness of the glove material and the nature of the task to be performed, gloves may substantially reduce tactile sensation and make it difficult or impossible to perform fine work. In spite of these disadvantages, gloves remain the most versatile approach to hand protection in the workplace.

Gloves may be made by molding a single thin film of the material to the shape of a hand. More commonly they are built with several layers, reinforcing areas of grip or palm contact, which may be textured for greater friction. A common practice is to impregnate cloth fabric with rubber or plastic and to use this as the foundation material for the gloves. Liners may be made of a water-absorbent material to keep the hand relatively dry for as long as possible, but the rate of wicking and evaporation of moisture away from the palm always seems to be less than the rate of perspiration.

Gloves come in many styles and models, and almost always in pairs. Selection of the right glove depends on the job to be done; the glove must be sufficiently flexible and strong, and must protect against exposure, provide adequate friction for a firm grip, and be as comfortable and dry as possible. The key factors to consider in selecting a glove for a particular purpose are therefore the composition of the material, the degree of protection, the finish on the palm and fingers or grip area, and the lining, if any. In general, there is a trade-off between tactile sensitivity through the material of the glove and dexterity, and the degree of protection afforded against organic solvents. A glove that provides greater dexterity but lesser protection against penetration may be used when contact with a chemical is likely to be infrequent and of short duration.

Glove materials are never absolutely impermeable; there is always the potential for small amounts of a chemical agent to get across given sufficient time. The degree to which a glove will resist penetration by a particular chemical or class of chemicals is the most critical issue in selecting the right product in a high-risk workplace. Glove manufacturers and safety equipment suppliers usually have charts or guides indicating the preferred glove types for handling particular chemicals. Several specific terms are in use in the testing reports provided by manufacturers to

describe the characteristics of glove materials. "Degradation" describes how long the glove will last in its intended use. "Permeation" describes the ease of penetration of a chemical through glove material, reported as the "breakthrough time" (BTT) from initial contact to detection of the material on the inside of the glove. It is also known as the "permeation rate at steady state," the maximum rate at which the chemical passes through the material expressed in mg/m<sup>2</sup>/sec. BTT is directly proportional to the thickness of the glove material; the permeation rate is inversely proportional to thickness. Gloves that do not provide enough protection to be used for long periods (such as natural rubber, for handling acetone) may be entirely adequate for protection against occasional splashes, when removed promptly.

With the selection of appropriate materials and the outside of the gloves washed periodically (being careful not to contaminate the inside), the use of gloves may provide excellent protection. To prevent the accumulation of chemicals, and to reduce the potential for inadvertent contamination of the inside of the glove (by handling the outside before putting it on), gloves made of impermeable materials should be washed with soap and hot water daily or after each work shift.

Latex, natural rubber, has many practical advantages when exposure is limited to dilute solutions, soapy water, or hot water. Latex can be extruded into very thin sheets, providing good dexterity and tactile sense, as in surgical gloves. The material is resistant to water, mild acids, mild alkalis, and alcohols. Gloves made out of natural rubber are soft, generally comfortable, and often have better dexterity than those made of other materials. Natural rubber seals well after puncture. However, latex has serious drawbacks. Natural rubber cannot be used for long periods under very hot conditions, nor can it be used for handling oil, grease, or solvents. Latex loses the property of resealing, and it stiffens and cracks with age and after exposure to oxidizing materials. The material melts with heat. Latex is easily punctured or torn, and dissolves readily in nonpolar organic solvents; it is therefore unreliable for most industrial purposes. Latex gloves are cheap, however, and work well for healthcare applications and for handling nonreactive powdered solids. The biggest drawback of latex is that it is a common skin sensitizer that triggers local dermatitis and some-

times severe systemic allergic reactions in sensitized users. Latex allergy is now a serious problem among healthcare professionals due to poor manufacturing practices when the demand of latex gloves soared, as universal precautions were being introduced. Talc and starch powder are used so that latex remains easily separated, and in the past the powder also contained free latex that could be inhaled.

Polyurethane gloves have characteristics of stretch, fit, and barrier protection similar to latex gloves. They are disposable and available as a non-allergenic alternative to latex gloves.

Polyvinyl chloride (PVC) materials are also tough and durable, but are generally stiff and do not stretch to conform to the hand. PVC gloves are resistant to water, acids, alkalis, alcohols, and most organic solvents, including chlorinated ones. PVC does not stand up well to heat, and is relatively easily damaged compared to other materials. PVC gloves are, depending on thickness and construction, more permeable than nitrile to some compounds, such as toluene, but less permeable to most. PVC gloves are inexpensive and non-allergenic. Although more expensive than latex, they are coming into common use in healthcare institutions.

Polyethylene gloves are unsatisfactory for most hard work, since their resistance to mechanical damage, heat, and oxidation is poor. They can be used for many of the applications mentioned above for latex gloves. They are inexpensive and useful when greater levels of protection are not needed. Chlorinated polyethylene, which is also inexpensive, has more desirable properties but still has less resistance to mechanical damage and oxidation than other materials.

The glove materials described above may be familiar to healthcare professionals. The glove materials described below are less often encountered in healthcare and more often seen in industry.

Synthetic rubbers (such as Neoprene®, a DuPont trade name) are highly resistant to water, acids, alkalis, alcohols, oils, organic solvents, and corrosives or other reactive chemicals. Gloves made of this material are very durable. They are commonly used in industry; in the models usually encountered, they tend to be somewhat stiff. Neoprene® resists oxidizing agents well; other synthetic rubbers may not. Neoprene® tends to be moderately expensive. Other synthetic rubbers, such as

styrene–butadiene, are less expensive but may not have as desirable characteristics for some applications. Butyl rubber is expensive and subject to mechanical damage, but is very resistant to heat and oxidation and to permeation by acetone and some other difficult chemicals.

Nitrile materials are very tough, flexible, and durable and resist heat better than natural or synthetic rubbers. Nitrile and nitrile mixes (including nitrile–butadiene) wear well but can be degraded by oxidizing agents; they are more resistant to oxidation than natural rubber but less so than Neoprene® or polyvinyl chloride. Nitrile gloves are highly impermeable to water, acids and alkalis, alcohols, and organic solvents, including many chlorinated hydrocarbons. They are moderately expensive. Thin nitrile gloves are also available for healthcare workers, including surgeons.

Certain compounds encountered in industry tend to degrade even nitrile and PVC gloves, and may break through synthetic rubber. These include chloroethane, 1,2-dichloroethane, dimethyl acetamide, ethylene dichloride, 1,1,1-trichloroethane, nitrobenzene, tetrahydrofuran, toluene, and trichloroethylene.

Leather gloves protect against laceration and abrasion. They should never be used in handling chemicals or very hot or cold materials because they absorb liquids and retain them in close contact with the skin.

Thermal insulating gloves may allow the wearer to grasp and carry materials under conditions of extreme cold or heat, but are too thick for other work. In the past, they were made of asbestos-containing materials, and significant exposure to asbestos resulted from their use.

Armored gloves or gauntlets, made of chain mail, are used for working with sharp instruments, for example in abattoirs, or for handling large animals that bite.

### **Other Protective Clothing**

Aprons protect the immediate “splash zone” of the chest and abdomen. Their major role is to prevent deposition on clothing that might be absorbed and transmitted to the skin. Protective aprons are made of the same general materials used for gloves, and the permeability-related considerations are similar.

Coveralls, laboratory coats, and smocks are intended to keep the wearer clean and to resist penetration by liquids. They also reduce the opportunity for contamination by low-level radioactive isotopes when these are handled. Cotton tends to be resistant to low-temperature ignition from small flames, as well. These clothing items are not intended to provide complete protection, however.

Polyethylene-coated garments are available to provide increased resistance to penetration by water and organic materials.

Workers in high-technology industries, healthcare facilities, or research institutions often like to wear clean, white lab coats as a sort of uniform outside the laboratory, hospital, or clean room, but wearing the same coat as worn at work should be discouraged. Workers should be encouraged to change into clean lab coats when leaving the area, and this should be mandatory if they have been working in an area where contamination may occur. Similarly, hospital workers wearing scrubs out into the community for short periods, such as lunch breaks, should be discouraged from doing so if they plan to return to the ward or surgical department.

Protective clothing should be kept and washed at the workplace and removed before leaving the workstation, to avoid spreading hazardous materials widely. Changing rooms or areas should be provided.

### **Eye Protection**

Eye protection is generally intended to protect the eye against traumatic hazards, dust, and chemical splashes, rather than light. Special lenses or barriers are used for protection against coherent light from lasers, ultraviolet and infrared radiation, as well as glare from intense light (see Chapter 9).

Protective clear glasses have hardened glass or plastic lenses (usually high-impact polycarbonate) to protect against direct injury from flying objects and to prevent glass from a shattered lens from injuring the eye. They also provide some protection against dust and wind. Clip-on models are available for wear over prescription glasses. Protective glasses can be ground to prescription, however, and this is

usually desirable for workers who require corrective lenses and who spend much time in hazardous workplaces. They are most effective when equipped with side shields, and this type should be used whenever there is a risk of flying objects or dust.

Protective lenses with filters and side shields protect against bright lights, infrared from hot objects, and ultraviolet radiation, such as reflections of welding flashes. They are not adequate protection against direct exposure to welding arcs or intense infrared or ultraviolet radiation.

Goggles may be plastic or glass, depending on the application. They should always be fitted with side shields. For prolonged use, the design should be resistant to fogging. Goggles are suitable protection against eye injury from flying objects, dust, chemical splashes, and molten metal, but they do not protect the rest of the face. Goggles are usually clear, but models are available with tinted lenses that provide modest protection against intense light, infrared, or ultraviolet radiation. Because they are applied to the face, goggles must be kept as clean as possible, and it is good practice to wash them in soap and hot water daily.

Face shields provide the greatest degree of protection for general purposes. Depending on the design, they protect against flying objects, chemical splashes, molten metal, infrared and ultraviolet radiation, and glare. They come in many models, including 180° clear shields that rest on the head comfortably like a headband. Others are mounted on hardhats, some of which incorporate earmuffs. Hand-held face shields can be used by occasional visitors or inspectors for temporary protection. Face shields, even those tinted for protection against welding flash, are not routinely used by welders, however, who must have both hands free and need more protection than most face shields afford.

Welders use helmets for most welding purposes; face shields may be used for inspection or observation at a distance, and goggles are common for certain purposes. Welding goggles are heavily tinted goggles that protect against mild infrared and ultraviolet radiation exposure and intense light. They provide minimal protection against flying objects and sputtering low-temperature molten metal, and are usually only used for brazing or soldering and for spot welding. They should

not be relied upon for eye protection in other applications. Welding helmets should always be used for general welding purposes.

The glass in the window of helmets or face shields, and the glass in goggles, is rated for protection against ultraviolet and infrared radiation. The rating should be checked before use to ensure that it is adequate for the type of welding undertaken.

Contact lenses do not provide substantial eye protection in industry, and appropriate protective equipment should always be worn over them.

## **Footwear**

Appropriate footwear is important for physical safety, particularly in workplaces where heavy or sharp objects might fall, where there are flying particles of hot or molten metal, or where wet or unsecured surfaces could result in slips or falls. Additionally, footwear that adequately supports the feet may reduce fatigue, back strain, and the risk of injuries to the feet, ankles, and knees.

Footwear for industrial use should provide a high level of traction for sure footing, provide adequate ventilation and keep the foot dry, resist puncture, and fit well with proper support. If intended for jobs in which sparking may be a hazard or in which static electricity is a problem, anti-static models are required; these are made of materials that are moderately conducting and dissipate the charge. If a hazard of electrocution exists in the workplace, electric shock-resistant footwear is required; these are made of insulating materials.

Work shoes and boots come in many varieties and with many options. Steel-toed shoes and boots are the standard for heavy industry, in order to prevent serious injury to the toes. Recently, thermo-plastic toe caps have been introduced that provide limited protection but do not conduct electricity or heat; these are more comfortable in extreme temperatures. Metatarsal guards are present on some models of work boots to protect against objects falling across the foot; external guards are generally preferred, but some models have the guard inside the boot.

Within the constraints of work requirements, workers should be encouraged to wear the most comfortable and supportive footwear available that is compatible with their duties. Safe footwear for

industrial use is now available in a wide range of fashions and colors. Workers in areas where there are no special hazards should be allowed to wear casual shoes unless footwear is a part of the job. It has become very common in New York and other urban centers for office workers to wear running shoes at work, and this trend has probably prevented a great deal of avoidable discomfort and lost productivity due to fatigue and back pain.

## **ENTERPRISE RISK ASSESSMENT AND MANAGEMENT**

This section presents and, in a sense, redefines risk assessment on two levels of application: enterprise risk assessment, meaning a systematic approach to identifying and controlling hazards in the company or organization wherever they occur; and workplace risk assessment, meaning identifying and controlling hazards at a particular workplace. This approach fits well with business management models of quality improvement and loss prevention.

“Risk assessment” and “risk management” have similar but somewhat different meanings in different contexts, such as in regulatory policy and standards setting and in loss prevention and liability management. In recent years the concepts of risk assessment and risk management have been adapted to occupational health at the enterprise and workplace level, principally by the United Kingdom’s Health and Safety Executive (HSE) and by occupational (industrial) hygiene organizations such as the International Occupational Hygiene Association and the American Industrial Hygiene Association. This has resulted in a terminology that uses many of the same words as policy-based risk assessment but with slightly different meanings and emphases when applied to occupational health and safety. As in policy-based risk assessment, a hazard is something that has the capacity to cause harm, whether it is a chemical or a condition in the workplace. A risk is the probability of something bad happening, such as an injury or health effect. A consequence is the result, the occurrence of that bad outcome. Obviously, a hazard, or threat, causes a risk of injury or disease only when a person is exposed to it.

## Enterprise Risk Assessment/Management and the Deming Cycle

In occupational health and safety, little distinction is made between risk assessment and risk management. Instead, the focus is on management principles, such as the Deming Cycle of quality improvement. The Deming Cycle is an approach developed by W. Edwards Deming, an American pioneer in quality improvement, whose ideas were enthusiastically adopted by Japanese industry in the 1950s and led to that country's industrial transformation from a low-cost producer to a high-value-added innovator. Deming sought a simple means of operationalizing for managers the somewhat inaccessible ideas of Walter A. Shewhart, who developed the basic theory of statistical quality monitoring in the 1930s. He came up with a radically simplified management system with four essential steps: Plan, Do, Study, Act (PDSA). "Plan" refers to conceptualizing from basics what improvements could be made and how to implement them. "Do" refers to putting the plans into practice. "Study" (the Japanese substituted "Check," but Deming strongly preferred "Study") refers to evaluation. "Act" refers to a correction to make the system work better. The cycle then repeats with a return to basics. The "Six Sigma" program, developed at General Electric, is a modern adaptation of the Deming Cycle.

Enterprise and workplace risk assessment and management are bound in a continuous Deming Cycle. This approach emphasizes technical solutions: avoidance, control or reduction of the hazard toward the prevention or minimization of risk. It makes hazard control accessible to managers and supervisors and equips them with a framework for solving many of their workplace hazard problems themselves. It lends itself particularly well to the control banding approach, which is described in the next section.

Hazard identification is the step in which a hazard comes to the attention of management or is recognized as a threat to health from research studies or the diffusion of information on health risks. Once a hazard is recognized, there is an obligation to manage it as a duty of corporate responsibility. Hazard identification initiates the process and commits the organization to begin the cycle.

Risk assessment at the enterprise level is the step in which the hazard is evaluated and a decision is made as to whether it is a big risk requiring high priority or a medium or small risk that can be prioritized among other business priorities and wait its turn for management attention. It also includes the step of weighing the options regarding what can be done to reduce or control the risk. Risk assessment in this context corresponds to the “Plan” phase of the Plan → Do → Study → Act cycle.

Risk management is a series of steps in which something is done about the risk, initially corresponding to the “Do” phase of the Plan → Do → Study → Act cycle. Risk management involves the following steps in sequence:

1. Risk reduction and control (“Do”)
2. Risk management through the life cycle of a product or facility (“Do”)
3. Risk management documentation (“Study”)
4. Formulation of risk management indicators (“Study”)
5. Review of experience (“Study”)
6. Correction and modification (“Act”)

Risk reduction and control may involve one or more of the workplace interventions described in Chapter 7. Risks can be terminated, treated, tolerated, or transferred to someone else (such as a contractor, or a hazardous function can be outsourced), but they are best avoided altogether if a hazard can be dispensed with, replaced, or removed.

Risk management through the life cycle refers to a comprehensive approach of managing risks from the initial step in the business and production cycle to the very end, when the product is delivered. In some industries, particularly the chemical industry, risk management continues and is managed through various programs (such as Responsible Care® in the chemical industry). Table 14.5 outlines two life cycles. One is for facilities; the other is for the product life cycle for the oil and gas industry, which happens to be open rather than closed because there is no recycling or re-use, at least of product for energy consumption.

**Table 14.5.** Risk Management Through Life Cycles: Oil and Gas Industry

Facilities	Products
Design	Exploration
Construction	Extraction
Commissioning	Processing
Operations	Refining
Support	Transportation
Maintenance	Marketing
Design	Distribution
Construction	Consumption

Risk management documentation refers to recording what was done and why, and reporting it to management. It is essential that there be a record of changes in the workplace so that in the future risks can be evaluated correctly and so that the results of the intervention can be studied and evaluated as the basis for further improvement. It is the data capturing phase of the “Study” step in the Plan → Do → Study → Act cycle.

Risk management indicators are performance indicators that inform management on the current level of risk in the workplace and on the results of interventions that have been made. At their simplest, they may be measured concentrations of contaminants in air, a leading indicator, or the frequency of work-related injury in a workplace, a trailing indicator. It is the information gathering phase of the “Study” step in the Plan → Do → Study → Act cycle.

Review of experience is the step where the information gained from indicators and the knowledge of the workplace captured by documentation are examined to determine if the level of protection has improved or is deficient and if specific interventions have reduced the risk or are not working. It is the analytical phase of the “Study” step in the Plan → Do → Study → Act cycle.

“Correct and adapt” is the step where mid-course corrections are made and the workplace is improved based on the information that has been captured and analyzed. For example, it may be determined

that a new ventilation system has solved the problem of exposure to a contaminant and nothing further is needed. However, if the ventilation system did not work, the action needed might be to use personal respiratory protection as a backup for health protection until the system can be fixed.

## Workplace Risk Assessment/Management

Risk assessment and management in the specific workplace brings the familiar cycle of “Plan → Do → Study → Act” down to its most basic level of practice. The approach outlined here was pioneered by the HSE as a means of simplifying hazard control in order to encourage and support good occupational health and safety practice for employers with limited resources and to reduce the need for expensive consultants. Equally important, it was designed to help managers determine when expensive consultants are actually needed.

After the hazard has been identified, it has to be evaluated. In occupational health and safety, this is done by determining how serious a risk it poses to employees, contractors, or other people who are likely to be exposed. There are three pieces of information that are most important:

- *Exposure opportunity:* How likely is it that people will come into contact with the hazard (most often by inhaling it) at levels that could be harmful? (One can call this level “over-exposure” for simplicity, because exposure at low levels is not the concern.)
- *Quantity:* How much of the hazardous material is used or is present in the workplace (grams, kilograms, pounds, tons)?
- *Level of hazard:* How dangerous is the hazard?

The level of hazard is usually determined by comparing the occupational exposure level (OEL) or other standard, by consulting standard references, or by referring to the manufacturer’s information on the material safety data sheet (MSDS).

Some risks are complicated and involve multiple ways in which things can go wrong. For example, an upset in a refinery is much more complicated than a welding accident. These complicated risk profiles usually require professional expertise. They can be modeled using conventional methods derived from engineering:

- Fault tree analysis is an evaluation based on flow charts showing how an adverse event (such as a fire, a blowout, or a spill) could happen. Fault trees work backward to determine what failure could cause the event, then what factors could cause the failure, and so forth, until a clear picture emerges of what would have to go wrong to result in a serious event. In many cases, more than one failure is required for an adverse outcome.
- HAZOP (hazard and operability analysis) studies are used to predict what could go wrong when a part or a process fails. They emphasize consequences and are particularly useful for complicated situations where there may be unanticipated consequences.

These advanced methods are not required for most risk assessment activities in the workplace. For most, simple methods such as checklists, grids, and physical inspection are enough.

With this information, risks can be classified and prioritized in two dimensions (see Table 14.6):

- How likely it is that an incident resulting in (over)exposure will happen? This information can be drawn from historical data, assuming that the workplace has not changed much. Otherwise, assumptions will have to be made.
- How severe will the consequences be if an (over)exposure occurs?

One tool for estimating how severe the consequences of an event might be is given in Table 14.7. It is only a general guide to consequence analysis. Hazards that cause little harm are given lesser priority,

**Table 14.6.** Classification of Risks

			Magnitude of Consequence		
		Low	Medium	High	
Probability of Occurrence	Medium	Medium priority	High priority	Highest priority*	
	High	Low priority	Medium priority	High priority	
	Low	No priority	Low priority	Medium to high priority	

\* These risks should not exist in a well-run operation. If they are present and uncontrolled, something is wrong.

but if a hazard has the potential to cause serious harm it must be attended to even if the likelihood (probability) of harm is low.

Hazards that result in risks that are both common and severe have usually been dealt with previously in the history of the enterprise. The principal exception is vehicular traffic accidents on the open road, which can be very serious and are the single leading cause of serious occupational injuries. Hazards that have the potential for catastrophic damage or intolerable harm should not exist in the normal workplace and their presence demands an explanation. Most remaining hazards in this priority category are an unavoidable part of operations in a critical industry, such as hydrogen sulfide in the oil and gas industry, for which protection is still possible through emergency measures. If a hazard in this category is not an unavoidable part of operations in a critical industry, then the manager must answer the question of why it was allowed to exist in the first place

Next highest in priority are hazards that may result in risk of severe damage or injury but are uncommon or occur rarely. These hazards still need to be controlled urgently, before an incident does occur.

High- and middle-priority hazards need to be reduced and controlled before they hurt someone or damage company assets. This

**Table 14.7** A Tool for Consequence Analysis (for a Well Site in the Oil and Gas Industry)

People	Building	Equipment	Environment	Reputation	Relationships	Production	Revenues	Wastage
Fire								
Explosion								
Blowout								
Spill								
Gas release								
Structural failure								
Cable snapping								
Theft								
Etc.								

For each column, rate the worst case for risks using the following scale (10 for catastrophic risks, 1 to 5 for all other, using the illustrations below for selected categories):

10 = fatality, total loss of significant assets, shutdown in production

5 = disabling injury, major property loss, prolonged interruption in production

4 = serious injury, significant property loss, interruption in production

3 = injury, reparable damage, one-day interruption in production

2 = minor injury, minor damage, brief interruption

1 = any injury, damage, interruption to production

0 = no impact whatever

---

Note: This grid does not imply that the outcomes are comparable between columns.

generally means making a special effort and not waiting until the next maintenance or replacement cycle. For example, if a relief valve is malfunctioning, it should be fixed without waiting for a plant shutdown. One should beware of risks that are relatively unlikely but that can cause serious harm or damage; they are usually underestimated. Because of their repetition, these “medium consequence” risks lead to a cumulative risk of injury over time, accumulated losses over time, and inefficiencies that translate into lost productivity—all of which are avoidable. Exposure to noise without hearing protection is an example of “medium consequence” risk severe enough to lead to disability (from deafness) over time. Such problems should be taken care of at the earliest opportunity. Managers should not be complacent about any risk that can cause serious harm, no matter how remote the possibility of the outcome.

Low-priority hazards may be inconvenient but only cause minor and tolerable risk, and therefore they can usually wait until the next maintenance or replacement cycle. Common risks with less severe consequences may still cause significant avoidable loss over time. Many of these hazards, such as vibration or low-grade hum in machinery that needs damping or poor ergonomic design (in the absence of a complaint or symptoms, which would raise the priority), are likely to increase the efficiency of the workplace and reduce distraction and so are worth doing even though they are of low priority.

Hazards that are unlikely to occur and that result in little risk of harm or loss are not important and are given no priority. They should be managed as part of upgrades during a plant shutdown or through periodic maintenance on a convenient schedule. For example, asbestos lagging (insulation) can be left in place as long as it is intact and not disintegrating, in order to prevent inadvertent release of fibers. Asbestos should be removed only when it is possible to seal off the area and use protective measures. This is usually best done during the time a plant is shut down for overhaul or renovation, when the area will be disrupted anyway.

This is a very general approach. The technical details will be different for each problem. Control banding, discussed in the next

section, represents a robust and straightforward way for managers and supervisors to deal with many of these problems themselves.

## **CONTROL BANDING**

Decisions on the appropriate intervention strategy in the workplace are normally the domain of occupational hygienists, who are guided by environmental monitoring. Normally, this would require extensive evaluation and measurement. However, it is clear that extensive environmental surveys are expensive and time consuming, and may be beyond the resources of small enterprises or enterprises in economically less developed countries. For occupational health protection to be brought to all workers, it is essential that it be accessible.

International bodies, led by the International Labour Organization and the International Occupational Hygiene Association, are promoting a new approach called “control banding” as a qualitative, less data-intensive approach to selecting appropriate interventions for worker protection. This approach has been refined by the Health and Safety Executive, the occupational health regulatory authority in the United Kingdom, and has resulted in a “control banding” system called Control of Occupational Substance Hazardous to Health (COSHH).

Control banding recognizes that in a typical workplace groups of workers have the same or similar exposure to various hazards. These groups are called “similar exposure groups” (SEG), and their common levels and patterns of exposure are collectively called an “exposure scenario” (ES). The groups are assembled by taking into account whether they use certain chemicals (without worrying about quantity at this stage), the department where they are assigned, and the job they are doing. A “job exposure matrix” (see Chapter 4) is used to estimate the most likely or the highest exposure levels associated with each SEG in normal operation.

Once the exposure level is approximately known, the hazard it may present is correlated with another source of information in the descriptors used by manufacturers and distributors of chemicals.

Chemical hazards are classified in one of four “bands” by their characteristics as described in the MSDS or product label, and by the estimated level of exposure.

For example, for planning respiratory protection, the bands might correspond to the following descriptors in literature from the manufacturer or from some authoritative body such as the International Programme on Chemical Substances (IPCS):

1. “Skin and eye irritants,” with exposure up to 500 ppm for a gas and 10 mg/m<sup>3</sup> for a dust (relatively low risk to lung but having the potential for irritation)
2. “Harmful on single exposure,” with exposure up to 50 ppm for a gas and 1 mg/m<sup>3</sup> for a dust (potentially toxic and likely to cause symptoms if exposure is not controlled)
3. “Severely irritating and corrosive,” with exposure up to 5 ppm for a gas and 0.1 mg/m<sup>3</sup> for a dust (likely to pose a threat of serious lung and mucosal irritation)
4. “Very toxic”/“reproductive hazard”/“sensitizer,” with exposure up to 0.5 ppm for a gas and 0.01 mg/m<sup>3</sup> for a dust (recognized as significantly toxic and requiring highly effective control measures)

For a given control strategy, continuing with the example of respiratory protection, each band corresponds to a generic level of control technology based whenever possible on standard equipment and methods. For example:

1. General area ventilation, housekeeping, and basic occupational hygiene
2. Local exhaust devices
3. Process enclosure
4. Specialized control measures as advised by a professional hygienist.

In this example, bands 1, 2, and 3 are progressively increasing levels of control, from ventilating the entire workplace, to removing

exposure by local exhaust at the point of origin, to separating the worker from the process. Band 4 triggers the decision to return to specific control measures adapted to the hazard, implying that for this band the generic approach will not work.

Similar control bands would be developed for skin protection and other applications.

The great advantage of control banding is that it moves directly to control without expensive area monitoring and surveys. Because it is based on standard methods, rather than controls designed for particular hazards, off-the-shelf technology can be used and costs contained. The disadvantage is that because no measurements are made, the process could fail and nobody would know it. However, the downside risk is managed by limiting the highest level of possible worker exposure to the more hazardous materials under the worst-case scenario.

Control banding holds great promise for expanding occupational health protection. It is not a panacea, but a way to solve 80 percent of the problem. This means that efforts and resources can be devoted to the remaining 20 percent, metaphorically.

## **RESOURCES**

- Levy BS, Wegman D. *Occupational Health: Recognizing and Preventing Work-Related Disease and Injury*. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2000.  
McLellan RK. *Guide to the Medical Evaluation for Respirator Use*. Beverley Farms, MA: OEM Press; 2000.

## **NOTEWORTHY READINGS**

- American Industrial Hygiene Association. *Noise and Hearing Conservation Manual*. American Industrial Hygiene Association. [Use current edition.]  
American Institute of Chemical Engineers, Institute for Chemical Process Safety. *Guidelines for Technical Management of Chemical Process Safety*. New York: Center for Chemical Process Safety; 1989.  
Easley CB, Laughlin JM, Gold RE, Tupy DR. Methyl parathion removal from denim fabrics. *Bull Environ Contam Toxicol*. 1981;27:101–108.  
Easter E. Removal of pesticide residues from fabrics by laundering. *Bull Environ Contam Toxicol*. 1983;15:29–33.  
Harber P. Medical evaluation for respirator use. *J Occup Med*. 1984;26:496–502.

## The Praeger Handbook of Occupational and Environmental Medicine

- Harber P, Balmes JR. Prevention of pulmonary disease in the workplace. *Occup Med.* 1991;6(1):1-152.
- Laughlin JM, Easley CB, Gold RE, Tupy D. Methyl parathion transfer from contaminated fabrics to subsequent laundry and to laundry equipment. *Bull Environ Contam Toxicol.* 1981;27:518-523.
- Olishfski JB, ed. *Fundamentals of Industrial Hygiene*. Chicago: National Safety Council; 1982. [Use current edition.]
- Performance of Protective Clothing*. ASTM STP 900. Philadelphia: American Society for Testing Materials; 1986.
- Raheel M. Resistance of selected textiles to pesticide penetration. *J Environ Health.* 1987;49:214-219.

# **15 CLINICAL EVALUATION**

The role of clinical evaluation in occupational and environmental medicine is more complex than simple case-finding. Occupational and environmental health problems occur with increased frequency in an exposed group, but not predictably in any one individual. Clinical medicine by its nature emphasizes the individual case. The result is great uncertainty when findings for groups of people are interpreted as they apply to individuals.

Despite these limitations, the role of medicine is not entirely one of response to individual illness in a setting of uncertainty. Carefully planned clinical protocols applied to a series of patients can convert data on individuals to meaningful descriptions of groups. Individual cases may be the starting point in new medical investigations, suggesting by their features or, more often, by their paradoxes that further investigation is needed to explain the observations adequately.

Clinical evaluation is a vehicle for prevention as much as for diagnosis and treatment in occupational and environmental medicine (OEM). Prevention of work-related illness for the individual, prevention of work-related illness for the working population, and prevention, to the extent possible, of illness due to personal and lifestyle factors are all part of clinical practice in OEM.

Clinical medicine is also a vehicle for public education and understanding. The public looks to physicians for guidance and reassurance about health matters with respect to the environment, and workers look for guidance on occupational risks. The OEM physician should be skilled in communication about occupational and environmental risks (see Chapter 7).

## CLINICAL ENCOUNTERS

Several distinct types of clinical encounters are common in occupational medicine. This chapter emphasizes the evaluation of occupational disorders:

- Pre-placement evaluations. Pre-placement evaluations are single-point screening studies performed to ensure that the worker's capabilities match the work to be done, that the worker is not extremely susceptible to harm from the work, and that he or she is not likely to pose a threat to others by being unable to work safely. They are performed after a decision has been made to hire the worker and an offer is made, but before the worker is assigned to the job. If a worker has a disability, the pre-placement evaluation will assess it. This evaluation will determine whether the disability is compatible with the tasks of the job, as well as what accommodations (either in general or under the Americans with Disabilities Act), if any, are required. Major issues associated with pre-placement evaluations include the validity of the test as an accurate predictor of work and safety performance, and the access of an employer to information on an employee's personal health. Pre-placement evaluations are discussed in Chapter 18, as part of the fitness-to-work evaluation.
- Pre-employment examinations (obsolete in the United States and Canada). Pre-employment examinations are evaluations undertaken before a decision is made to hire the worker and before an offer is made. They are performed when the employer requires good health or a normal physical examination as a

requirement for the job, rather than the capacity to perform specific job-related tasks. Pre-employment examinations are not in keeping with the Americans with Disabilities Act (ADA) and the Canadian Charter, and can get the physician and the employer in trouble. They are mentioned here as a warning: if an employer in North America uses the terminology “pre-employment,” that employer is very likely to have regressive human resources policies and to be insensitive to legal requirements. The intricacies of this are explained in Chapter 18. Pre-placement examinations are still used in other parts of the world.

- Fitness for return to work. These evaluations are similar to pre-placement evaluations except that they pertain to a worker’s ability to carry on following an injury or illness rather than to his or her ability to begin work. A worker who is returning to a job should generally be held to the same criteria as a new worker, but with due consideration for having successfully done the job before. The evaluations are also discussed in Chapter 18.
- Disability (or, more properly “impairment”) evaluations. These are medical evaluations intended to document functional impairment for use in determining total disability, usually for eligibility for compensation (see Chapter 18).
- Occupational health monitoring. Monitoring is the systematic search over time for health problems of many types, which may or may not be related to workplace exposure. It is a way of detecting new occupational disorders and identifying previously unsuspected risk factors. These evaluations are discussed in Chapter 6.
- Periodic health surveillance. Surveillance is different from monitoring because one is looking for a specific health outcome that is known in advance. The purpose of surveillance is to determine whether a problem continues to exist despite measures to control it. The potential for labeling exists if workers are led to believe that because of their exposure to a hazard their health is likely to suffer. These evaluations are discussed in Chapter 6.

- Case finding for occupational disorders. When a suspected occupational disorder is identified, the case should be as thoroughly documented as possible. This information will be used by both sides in a workers' compensation case, and may be critical to resolving the problem and preventing future exposure.
- Case finding for health-related risks. In the routine evaluation of a personal health problem or a complaint that is not initially suspected to be work-related, an occupational disorder may be identified. For example, the opinion of an OEM physician is often sought in cases of peripheral neuropathy, because of the associations of many occupational exposures with the condition.
- Independent medical evaluation (IME). OEM physicians, as well as other physicians besides the treating physician in the case, may be asked to conduct a thorough or a narrowly focused evaluation of a case in order to ascertain issues of fact, opinions about causation, or ratings of the degree of impairment in a disputed case. The IME may be requested by either side in a legal action, by a workers' compensation carrier or board, by an insurance company, or by order of a judge. IMEs are unusual in that there is no real or implied physician-patient relationship. The injured worker is called an "examinee" and the party who requested the IME is called the "client"; neither is a patient. The physician is not obligated or expected to describe his or her findings to the examinee, to recommend treatment, to make a referral, or to offer an opinion at the time of the examination. There is an ethical obligation to inform the examinee if something imminently threatening is found on the examination. The physician should not treat the condition or attempt to recruit the examinee as a patient. In California, IMEs are known as "agreed medical examinations." They are sometimes called "impartial," "neutral," or "binding" medical evaluations, the latter implying that the physician performing the IME is either breaking a tie with respect to experts for the plaintiff and defense, or will be considered the accepted authority.
- Occupational and environmental health research.

Some disorders are not due to a single incident but arise out of chronic or intermittent exposure to a hazard over time. Injuries to intrinsically weak body parts due to repeated movements, noise-induced hearing loss, vibration-induced vasospasm ("white finger"), and occupational cancer are examples of disorders which may occur over time after an individual worker has sustained repeated exposure in similar jobs for two or more employers. Such disorders are called "cumulative injuries" in workers' compensation terminology, or "repetitive strain injuries" for musculoskeletal conditions in OEM, because they are theoretically an outcome of cumulative exposures over time, as if each exposure caused a tiny injury, which injuries in sum produced the disorder. In such circumstances, a comprehensive occupational and environmental history, incorporating a detailed work history with as much information on sources of exposure as can be obtained, helps the OEM physician understand the magnitude of the ergonomic strain. It also helps the workers' compensation board to apportion impairment based on the proportionate contribution of a known exposure during a term of employment to the final outcome. In diseases with a long latency, such as cancer, weight should be placed on earlier exposures.

Another complicated situation arises when second (or subsequent) injuries occur in the presence of an existing impairment, such as re-injury of an already injured back, or a sprain in a hand that has had a finger amputated. Workers' compensation boards have special rules for dealing with second injuries, usually by apportioning the total degree of impairment between the first and second injuries.

## **THE OCCUPATIONAL AND ENVIRONMENTAL HISTORY**

The occupational and environmental history is as much a part of a thorough medical history as are the family and past medical history. The ability to take such a history is as much a part of OEM practice as performing a neurological examination is of practicing neurology.

Like every other component of a medical interview, an occupational and environmental history can be abbreviated, but it should never be omitted in a thorough evaluation. Unfortunately, in the absence of training and understanding on the part of physicians regarding toxicology and occupational exposure, the occupational and environmental history may be collected but not understood. Without a means of easy interpretation, the occupational and environmental history is reduced to a bewildering catalogue of unfamiliar chemicals, physical hazards, and jobs. The physician who has a responsibility for and an interest in the identification and documentation of occupational disorders should develop a systematic approach to the occupational and environmental history, and become familiar with the sources of information needed to interpret it.

### **Levels of the Occupational and Environmental History**

The occupational and environmental history can be collected on four levels (Table 15.1).

**Table 15.1.** Levels and Applications of the Occupational and Environmental History

Level	Description	Used in Evaluation of:
Basic	Current occupation or employment status	Acute disorders, future disease risk, fitness to work
Diagnostic	Designed to rule in or out suspected association	Acute or chronic disorders
Screening	Complete database for occupations and known exposures	Surveillance
Comprehensive	Comprehensive and detailed profile of work-, avocation-, and environment-related exposures	Complex problems

### ***Basic***

In the simplest cases, including most acute injuries, the physician only needs to know the patient's current occupation, or—if the patient is not working—whether the patient is unemployed or retired. Even in simple cases, knowledge of the patient's occupation helps in patient management. It addresses three common concerns: (1) Is there a direct occupational association that bears on the patient's condition and that may alter priorities in management? For example, does a patient's occupation or aspirations require that particular attention be placed on rehabilitation of fine motor skills (such as for a musician or surgeon), maintenance of a regimen to maintain overall strength during recovery (such as for an athlete or dancer), or unusual cosmetic concerns (such as for a model or actor). (2) Is there an occupational association reportable by law? (3) What are the implications for the patient's return to work?

### ***Diagnostic***

More complex medical problems may require questions intended to identify a specific suspected or plausible association, exploring the patient's recent or past job assignments or work practices, but not necessarily covering the entire occupational and environmental history. At this level, particular associations are sought, and the occupational and environmental history may provide confirmation. For example, a case of aplastic anemia may be investigated in some detail to identify potential sources of exposure to benzene, but little attention may be given to unrelated parts of the work history.

### ***Screening***

A more complete occupational and environmental history is used for screening individuals or groups. Usually, this level of the occupational and environmental history is used either for surveillance of patients at risk for certain conditions, who are followed for early detection of those conditions, or for periodic health screening of

individuals in generally good health, to prevent disease or common illnesses at an early stage of development. The occupational and environmental history at this level is usually taken in some detail at the beginning of the program, and then supplemented or reported annually or at least periodically. Usually, the interim history highlights significant exposures, provides a database for adapting the screening procedure if necessary, and, together with the baseline history, allows reconstruction of the complete history of exposure should an occupation-related condition develop later. Periodic updating of the history supplements environmental exposure measurement in monitoring the worker's potential for exposure to occupational hazards and is a necessary part of the surveillance of workers at high risk, especially those who change jobs often. This documentation protects employers against liability for problems associated with exposure on previous jobs, and assists employees in supporting a valid claim by providing an objective record of exposure.

### *Comprehensive*

The comprehensive occupational and environmental history is performed as part of a thorough investigation of complex medical problems. At this level the history is complete and detailed, and also covers non-occupational risk factors. A comprehensive evaluation by the physician is impractical in some situations, and may require consultation with specially trained occupational health professionals in order to gain insight into the nature and effects of occupational exposures. Disorders commonly requiring this approach include interstitial fibrosis or pneumonitis, the pneumoconioses, peripheral neuropathies, bladder cancer, and leukemias. Such cases often require an evaluation far beyond the scope of the occupational and environmental history taken by the physician. An occupational and environmental history can rule out certain possibilities or suggest a possible diagnosis. For example, a patient with a peripheral neuropathy may report having been exposed to any of a variety of hazards that could cause this condition, such as lead and certain solvents. If the

occupational and environmental history suggests exposure to an occupational hazard, the worker may have a disease that is compensable under workers' compensation, and that is reportable by law.

Several questionnaires for recording the occupational and environmental history have been developed. Most have serious drawbacks, including insufficient interview questions, overly extensive detail on nonessential topics, poor format or flow, lack of internal cross-checks, and lack of integration with other parts of the medical record. The form for the comprehensive occupational and environmental history presented here is adapted from a widely used version as designed by the American Lung Association of San Diego and Imperial Counties, and published in the Annals of Internal Medicine in 1983 (Figures 15.1 and 15.2). This instrument was refined after extensive field trials. In this format, the questionnaire was designed to be part of a complete system and to strike a balance between completeness and ease of administration. No one questionnaire can be comprehensive for all possible occupational illnesses, but there should be sufficient detail to provide leads for follow-up. The form provided here is self-administered and can be mailed to the patient before appointments.

### **Using the Occupational and Environmental History**

Pre-mailing or e-mailing the occupational and environmental history, to be completed in advance of the appointment, saves time in the clinic and allows patients or their families to consult personal records such as income tax returns or union documents. This form is divided into three parts, each of which may consist of several sections and any number of questions (Table 15.2). The "occupational profile" is an inventory of occupational experience to examine the potential for past exposure to hazards. The inventory of "occupational exposure" is for the purposes of identifying susceptibility, obtaining a history of occupation-related disorders, and confirming the profile of exposure to hazards. The environmental history alerts the physician to exposures to hazards in the home or other places (not strictly occupational) that may be significant. Each of the three parts of an occupational and

# The Praeger Handbook of Occupational and Environmental Medicine

## Occupational Environmental History Form

### I IDENTIFICATION

Name: \_\_\_\_\_

Soc. Sec. \_\_\_\_\_

Address: \_\_\_\_\_

Sex: M F

Telephone: home \_\_\_\_\_ work \_\_\_\_\_

Birthday \_\_\_\_\_

### II OCCUPATIONAL PROFILE

Fill in the table below listing all jobs at which you have worked, including short-term, seasonal, and part-time employment. Start with your present job and go back to the first. Use additional paper if necessary.

Workplace (Employer's name and address or city)	Dates worked From      To	Did you work full time?	Type of Industry (Describe)	Describe your job duties	Known health hazards in workplace (dusts, solvents, etc.)	Protective equipment used?	Were you ever off work for a health problem or injury?

**Figure 15.1.** Recommended form for taking the occupational history (front side). Adapted from Guidotti TL, Chairman: Committee on Occupational and Environmental Health of the American Lung Association of San Diego and Imperial Counties. Taking the occupational history. *Ann Intern Med* 1983; 99: 641–51. Reprinted with permission.

environmental history must be completed in order for the evaluation to be complete. Omission of any section may result in critical information being left out, and intentional redundancy being lost.

Two particularly important columns in the occupational profile are headed “Description of Job Duties” and “Known Health Hazards.” The inventory of occupational exposure is arranged so that a quick glance identifies a positive response. Responses in this section should be cross-checked against the “Known Health Hazards” section of the

### III. OCCUPATIONAL EXPOSURE INVENTORY

1. Please describe any health problems or injuries you have experienced connected with your present or past jobs.
2. Have any of your co-workers also experienced health problems or injuries connected with the same jobs? No Yes  
If yes, please describe:
3. Do you or have you ever smoked cigarettes, cigars or pipes? No Yes  
If so, which and how many per day:
4. Do you smoke while on the job, as a general rule? No Yes
5. Do you have any allergies or allergic conditions? No Yes  
If so, please describe:
6. Have you ever worked with any substance which caused you to break out in a rash? No Yes  
If so, please describe your reaction and name the substance:
7. Have you ever been off work for more than a day because of an illness or injury related to work? No Yes  
If so, please describe:
8. Have you ever worked at a job which caused you trouble breathing, such as cough, shortness of wind, wheezing? No Yes  
If so, please describe:
9. Have you ever changed jobs or work assignments because of any health problems or injuries? No Yes  
If so, please describe:
10. Do you frequently experience pain or discomfort in your lower back or have you been under a doctor's care for back problems? No Yes  
If so, please describe:
11. Have you ever worked at a job or hobby in which you came into direct contact with any of the following substances by breathing, touching, or direct exposure? If so, please check the box beside the substance:

<input type="checkbox"/> Acids	<input type="checkbox"/> Beryllium	<input type="checkbox"/> Chromates	<input type="checkbox"/> Heat (severe)	<input type="checkbox"/> Nickel	<input type="checkbox"/> Radiation	<input type="checkbox"/> Trichloroethylene
<input type="checkbox"/> Alcohols	<input type="checkbox"/> Calcium	<input type="checkbox"/> Coal dust	<input type="checkbox"/> Isocyanates	<input type="checkbox"/> Noise (loud)	<input type="checkbox"/> Rock dust	<input type="checkbox"/> Trinitroluene
(Industrial)	<input type="checkbox"/> Carbon	<input type="checkbox"/> Cold (severe)	<input type="checkbox"/> Ketones	<input type="checkbox"/> PBBs	<input type="checkbox"/> Silica powder	<input type="checkbox"/> Vibration
<input type="checkbox"/> Alkalies	<input type="checkbox"/> Dichlorobenzene	<input type="checkbox"/> Lead	<input type="checkbox"/> PCBs	<input type="checkbox"/> Solvents	<input type="checkbox"/> Styrene	<input type="checkbox"/> Vinyl chloride
<input type="checkbox"/> Ammonia	<input type="checkbox"/> Chlorinated	<input type="checkbox"/> Ethylene dibromide	<input type="checkbox"/> Manganese	<input type="checkbox"/> Manganese	<input type="checkbox"/> Talc	<input type="checkbox"/> Welding fumes
<input type="checkbox"/> Arsenic	<input type="checkbox"/> naphthalenes	<input type="checkbox"/> Ethylene dichloride	<input type="checkbox"/> Mercury	<input type="checkbox"/> Pesticides	<input type="checkbox"/> Toluene	<input type="checkbox"/> X-rays
<input type="checkbox"/> Asbestos	<input type="checkbox"/> Chloroform	<input type="checkbox"/> Fiberglass	<input type="checkbox"/> Methylene	<input type="checkbox"/> Phenol	<input type="checkbox"/> TDI or MDI	
<input type="checkbox"/> Benzene	<input type="checkbox"/> Chloroprene	<input type="checkbox"/> Halothane	Chloride	<input type="checkbox"/> Phosgene		

If you have answered "yes" to any of the above, please describe your exposure on a separate sheet of paper.

### IV. ENVIRONMENTAL HISTORY

1. Have you ever changed your residence or home because of a health problem? No Yes  
If so, please describe:
2. Do you live next door to or very near an industrial plant? No Yes  
If so, please describe:
3. Do you have a hobby or craft which you do at home? No Yes  
If so, please describe:
4. Does your spouse or any other household member have contact with dusts or chemicals at work or during leisure activities? No Yes  
If so, please describe:
5. Do you use pesticides around your home or garden? No Yes  
If so, please describe:
6. Which of the following do you have in your home? (Please check those that apply) No Yes  
 Air conditioner    Air purifier    Humidifier    Gas stove    Electric stove    Fireplace    Central heating

**Figure 15.2.** Recommended form for taking the occupational history (reverse side). (Reproduced with permission of *Annals of Internal Medicine*.) Adapted from Guidotti TL, Chairman: Committee on Occupational and Environmental Health of the American Lung Association of San Diego and Imperial Counties. Taking the occupational history. *Ann Intern Med* 1983; 99: 641–51. Reprinted with permission.

**Table 15.2.** Structure of the Occupational History.

*Occupational Profile*

- Past occupational health problems
- Protective equipment, ventilation
- Known health hazards\*
- Job duties, occupational title
- Industry
- Dates, full- or part-time

*Inventory of Occupational Exposures*

- Health checklist
- Exposure checklist\*
- Known allergies, sensitization
- Smoking history

*Environmental History*

- Home environment
- Passive exposure
- Avocational exposure (hobbies, volunteer work)

\* Intentional duplication to ensure completeness and thorough reporting.

occupational profile. Redundancy is built into this form intentionally, to aid in detecting exposures that might otherwise be missed due to the wording of a question or poor recollection. Workers tend to remember and record on the form only the unusual health hazards associated with their workplace. They often take for granted exposures that are routine for them, forgetting that these exposures may be unfamiliar to the interviewer. Entries made in the column “Known Health Hazards” in an occupational and environmental history should therefore never be considered as a complete inventory of exposure.

Establishing an occupational association requires an occupational and environmental history and a complete inventory of exposure at work, including the constituents of proprietary compounds. Tables 15.3 and 15.4 present inventories of common occupational exposures likely to be encountered in practice. Neither list is intended to be comprehensive.

The job duties should be interpreted in light of the industry in which the patient is employed. The job descriptions listed in the

**Table 15.3.** Exposure Inventory for the Occupational and Environmental History

---

*A. Aerosols, Irritants, and Gases*

- a. Carbon monoxide
- b. Ethylene oxide
- c. Formaldehyde
- d. Inert gases
- e. Hydrogen sulfide
- f. Nitrogen dioxide
- g. Ozone
- h. Phosgene
- i. Sewer gas (predominantly hydrogen sulfide)
- j. Smoke
- k. Sulfur dioxide

*B. Biological Airborne Hazards*

- a. Bacteria
- b. Fungi
- c. Molds
- d. Organic dusts
- e. Spores
- f. Viruses
- g. Toxins

*C. Highly Reactive Substances*

- a. Acids
- b. Alkalies
- c. Amines
- d. Ammonia
- e. Chlorine
- f. Hydrazine
- g. Phenols

*D. Dyes and Stains*

- a. Aniline dyes
- b. Azo dyes
- c. Benzidine
- d. Solvents
- e. Other coatings and surface treatments

(Continued)

**Table 15.3. (Continued)**

---

*E. Inorganic Dusts and Powders*

- a. Asbestos
- b. Coal dust
- c. Fiberglass
- d. Man-made mineral fibers
- e. Silica
- f. Talc

*F. Pesticides*

- a. Carbamates
- b. Organochlorines
- c. Organophosphates
- d. Phenoxyherbicides
- e. Pyrethroids

*G. Radiation*

- a. Hospital and Healthcare
- b. Infrared
- c. Laser
- d. Microwaves
- e. Radioisotopic wastes
- f. Radionuclides, including radon
- g. Ultraviolet light
- h. X-rays

*H. Metals and Metal Fumes*

- a. Aluminum
- b. Arsenic, arsine
- c. Beryllium
- d. Cadmium
- e. Chromium
- f. Cobalt
- g. Iron
- h. Lead
- i. Mercury

*I. Organic Dusts*

- a. Cotton dust
- b. Poison oak
- c. Wood dusts

**Table 15.3. (Continued)**

---

*J. Petrochemicals*

- a. Asphalt and tar
- b. Creosote
- c. Coal tar
- d. Dioxins and furans
- e. PBBs (polybrominated biphenyls)
- f. PCBs (polychlorinated biphenyls)
- g. Petroleum distillates
- h. Hydrogen sulfide

*K. Physical Agents*

- a. Cold stress
- b. Heavy lifting
- c. Noise
- d. Thermal stress
- e. Vibration

*L. Plastics, Polymers, Composites, and Monomer Constituents*

- a. Acrylonitrile
- b. Aliphatic amines
- c. Epoxy resins
- d. Phthalates
- e. Styrene
- f. Toluene diisocyanate (TDI)
- g. Vinyl chloride

*M. Welding and Related Emissions* (includes many of the exposures listed under A, G, and H)

- a. Cadmium
- b. Copper
- c. Lead
- d. Nickel
- e. Nitrogen oxides
- f. Ozone (aluminum welding)
- g. Zinc

*N. Solvents*

- a. Benzene
- b. Methylene chloride
- c. Perchloroethylene ("perc")

(Continued)

**Table 15.3. (Continued)**

- 
- d. Toluene
  - e. Trichloroethane
  - f. Trichlorethylene (“trike”)
  - g. Xylene(s)
- 

**Table 15.4. Occupations and Exposures to Major Hazards**

---

Agriculture/farming/pest control	A, B, F, K, M
Automobile/aircraft mfg. and repair	A, C, E, H, K, M
Baking/food handling	B, L, M
Boiler operations and cleaning	A, C, E, K
Carpentry/woodworking/lumber industry	B, I, J, K, N
Ceramics and masonry	E, H
Chemical industry, biotechnology, and users of chemicals	A–N
Construction/demolition/road work/ maintenance/plastering	C, D, E, K, J, M, N
Dry cleaning/laundry	D, K, N
Electricity/electronics	A, C, E, H, J, N
Foundry work	A, C, E, H, K, M, N
Healthcare/laboratory work/dental work	A–E, G, J–L, N
Machinery/grinding/metalwork	A, C, H, K, M, N
Mining	A, E, G, K, M
Oil and gas/petrochemical industry	A, C, G, J, K, N
Paper industry	E, H, N
Plastics manufacturing/molding	E, J, L
Plumbing/pipefitting/shipfitting	A, C, E, H, K, M
Printing/lithography	D, K, N
Sandblasting/spray painting	A, E, H, K, N
Textile industry	A, D, K, I, N
Transportation maintenance (shipyard/dockyard)	A, C, E, H, J, K, N
Welding	A, E, H, M

---

patient's occupational profile should be interpreted with respect to those industry headings listed in Table 15.4. Many jobs that carry similar titles differ from industry to industry, such as "firemen," "engineers," and "mechanics." An "engineer" may work at a desk, in a plant, or in the cab of a train. The worker may not know all of the possible exposures that occur on the job. Additional information may be available from the Manufacturer's Safety Data Sheet (MSDS), a document provided by the manufacturer on request and usually kept in the workplace. Information may also be obtained from reference literature in medical, chemical, or engineering libraries, especially for common industrial processes. Additionally, it should be kept in mind that workers in certain trades, such as painting and welding, may be exposed to hazards from work done by co-workers nearby or by others in the same job. For example, electricians may be exposed to insulation materials being installed in the same room.

The duration of employment and the degree of personal protection are essential pieces of information. Duration is particularly important in determining whether enough time has elapsed to account for the latency period for the disease in question. Latency periods must be interpreted carefully, however, because there is much variability in individual cases. Mean latency periods of 20 years for most occupational cancers, for example, are not minimum intervals required before cancer can be attributed to exposure, but only a most probable, average interval and may be irrelevant to a given case. The occupational and environmental history can be applied to many purposes besides diagnosis and assessing the likelihood of an occupational cause for a condition (Table 15.5). A thorough occupational and environmental history is often decisive in establishing a worker's eligibility under workers' compensation. The profile of employment and skills can be important in the evaluation of disability and the implications of a given impairment for future employment. Employers can use the occupational and environmental history as a guide to monitoring the risks to which their employees are or have been exposed, and to document whether a problem was present prior to hiring the worker. An

**Table 15.5.** Uses of the Occupational and Environmental History

- 
- Diagnosis: identifying cause of disorder
  - Evaluating fitness to continue working in usual occupation
  - Surveillance of workers at high risk
  - Assessing eligibility for compensation
  - Assessing legal liability
  - Worker education
  - Community education
- 

occupational and environmental history may suggest a need for close monitoring if the worker has been exposed to a toxic substance, such as lead or asbestos, that may have long-term effects.

When the occupational and environmental history is used for diagnosis or for a comprehensive exploratory evaluation, additional interview questions are required. Some of the most useful follow-up interview items are listed in Table 15.6.

**Table 15.6.** Supplemental Questions to the Occupational and Environmental History

---

Question	Interpretation
Is your condition better or worse when you are off work for a few days or on vacation?	Identify patterns suggesting either improvement or exacerbation on withdrawal from exposure.
Is your condition better or worse when you return to work after a weekend or vacation?	Identify patterns suggesting return of condition on re-exposure in the workplace.
Does your condition get worse or better after you have been back at work for several days or shifts?	Identify patterns suggesting either tolerance or cumulative effects with multiple exposure.
Describe your workplace. (Please draw a diagram and indicate your work station.)	Evaluate proximity to exposure, protection available (e.g., ventilation or sound barriers), mobility within the workplace, location of co-workers who may also be affected. <i>(Continued)</i>

---

**Table 15.6.** (*Continued*)

What ventilation systems are used in your work space? Do they seem to work?	Obtain general impression of adequacy of ventilation, by movement of air and odors.
Does the protective equipment you are issued fit properly? Do you receive instructions in its proper use? Do you ever fix or make changes in the equipment to make it more comfortable?	Consider the possibility that protective equipment is not fully effective. In the case of respirators (masks), ask if they were “fit-tested” to comply with occupational health and safety regulations.
Where do you eat, smoke, and take your breaks when you are on the job?	Identify opportunities for food- and cigarette-borne intake, adequacy of rest stations (isolation from heat, noise, fumes, etc.).
Where are your (your spouse’s or partner’s) work clothes laundered?	Identify possibility of passive exposure at home or prolonged skin contact.
How often do you wash your hands at work, and how do you wash them?	Identify potential for contamination of hands or contact with solvents or drying agents.
What is your spouse’s or partner’s occupation?	Identify potential for passive exposure. Occupational and environmental history for partner may be indicated.
Have any of your fellow workers experienced similar conditions?	Identify others who may have been affected; may lead to inquiries that clarify the individual patient’s problem. Prevention-oriented interventions or requests for investigation by the appropriate occupational safety and health agency may be required.
Do you recall a specific incident or accident that occurred on the job? Were others also affected?	Identify unusual or transient conditions that may have resulted in an exposure not reflected in the occupational and environmental

(*Continued*)

**Table 15.6.** (Continued)

Are animals (pets, livestock, birds, or pests such as mice) present in the vicinity? Has there been a change in their health, appearance, or behavior?	history, such as leaks, fires, or uncontrolled exothermic chemical reactions.
Has there been a change in the process, job responsibilities, workplace configuration, or environment? When did the change occur in relation to the onset of the disorder?	Animals (and especially animal wastes) may be a source of infectious or antigenic hazards. Animals may also respond to toxic exposures that affect humans.
Do your co-workers smoke on the job? How intensively and for how long are you exposed to cigarette smoke?	Identify precipitating factors that led to exposure.
	Identify symptoms related to passive smoking (usually in non-smokers) or smoking interactions.

## **CLINICAL EVALUATION OF OCCUPATIONAL DISORDERS**

Recognition of occupational and environmental illnesses frequently demands a higher index of suspicion than is the case for non-work-related conditions. Occupational and environmental diseases are rarely recognized on sight. There are three essential mental steps:

1. Recognizing the illness,
2. Recognizing the likely cause, and
3. Recognizing the work situation in which the exposure occurred.

Most clinicians find that following a definite order of investigation prevents omissions. Table 15.7 presents the common order of clinical

**Table 15.7.** Clinical Assessment in Occupational Illness

Component of the Encounter	Significance in Occupational Health
1.0 Interview	Awareness of affect, demeanor.
1.1 Chief complaint	Consider occupational associations.
1.2 History of the present illness	Probe for occupational associations.
1.3 Past medical history	Past occupational illnesses noted, possible susceptibility to exposures.
1.4 Personal history	Patient's social situation.
1.5 Family history	Genetic susceptibility to exposures.
1.6 Occupational and environmental history	Review self-administered questionnaire and inquire further into possible associations.
1.7 Review of symptoms	Alertness to symptoms.
2.0 Physical examination	Alertness to signs.
3.0 Laboratory findings	Alertness to findings.
4.0 Radiological findings	Alertness to findings.

evaluation and points out the significance of each step in the recognition of occupational illness.

Integration of the occupational and environmental history with the clinical evaluation requires that three questions be answered in sequence:

1. To what occupational hazards might the patient have been exposed (agent)? (See Tables 15.3, 15.4, and 15.6.)
2. What factors inherent in the host may predispose or modify the response to an occupational hazard (susceptibility)? (See Table 15.8.)
3. What circumstances were present in this situation that allowed the exposure to occur and that may have modified the pattern of its presentation (environment)? (See Table 15.6.)

**Table 15.8.** Some Factors That May Modify Risk of Occupational Illness

Modifying Factor	Known or Probable Effect
<i>General</i>	
1. Age	Youth: latency may result in cancer later in life following exposure. Elderly: more susceptible to toxicity.
2. Sex	Sex differences exist for some toxicity states; reproductive effects.
3. Smoking status	Confers additive risk in some situations. Confers interactive risk in some situations.
3.1 Current smoker	
3.2 Smoker at time of exposure	
3.3 Smoking during exposure	Modifies toxic exposure in some situations (e.g., polymer fume fever).
4. Family history	Hereditary conditions or predispositions may be exacerbated or triggered (e.g., cancer-prone families).
5. Exercise	
5.1 Conditioning	Fitness may reduce susceptibility in some situations.
5.2 At time of exposure	Generally, increased susceptibility.
6. Metabolic states	Activity of certain enzyme systems involved in activation, detoxification, and adaptation to toxic exposures may modify response, although remaining within the normal range of enzyme activity.
<i>Medical</i>	
1. Atopy	Usually, any debilitating condition can enhance clinical susceptibility.
1.1 Asthma	Increased bronchial reactivity.
1.2 Eczema	Tendency toward easy sensitization.
2. Chronic respiratory disease	Increased bronchial reactivity.

(Continued)

**Table 15.8.** (*Continued*)

2.1	Respiratory insufficiency	Diminished pulmonary reserve.
2.2	Bronchitis	Exacerbated bronchial irritation.
3.	Chronic cardiovascular disease	Increased susceptibility.
3.1	Cardiac insufficiency	Angina in some situations (e.g., with carbon monoxide, methylene chloride).
3.2	Coronary artery disease	
4.	Infection	
4.1	Acute viral illness	Increased susceptibility to bronchial irritation, possibly synergistic effects.
4.2	Exposure to infectious agents	Certain exposures may depress host defenses.
5.	Immune deficiency states	Immunodeficiency; increased susceptibility to toxic effects in some situations.
5.1	Hereditary	
5.2	Immunosuppressive therapy	
5.3	AIDS	
6.	Renal disease	Increased susceptibility to infections.
6.1	Renal insufficiency	Additive or synergistic effects may occur with exposure to nephrotoxic agents.
6.2	Chronic renal disease	Increased susceptibility to toxic agents excreted via renal route.
7.	Neurologic conditions	
7.1	Diminished mental capacity	May affect judgment and response to exposure.
7.2	Neurologic disease	Toxic effects may be additive; increased clinical susceptibility.
7.3	Seizure disorder	Certain toxic exposures may alter threshold.
7.4	Impaired perception (visual, olfactory, auditory)	Impaired ability to avoid hazard.

*(Continued)*

**Table 15.8. (Continued)**


---

8. Dermatologic conditions	Rashes may increase dermal absorption, may aggravate occupational dermatitis, or may condition response.
9. Substance abuse	Alcohol and drug abuse may have additive or interactive effects. Many interactions occur with smoking. Both may mask or mimic effects of occupational exposure.
9.1 Alcohol	
9.2 Smoking	
9.3 Drugs	
10. Hepatic insufficiency	Increased susceptibility to toxic agents detoxified by liver; increased susceptibility to hepatotoxic agents; reduced hepatic reserve.
11. Systemic conditions	Increased susceptibility to toxic effects.
11.1 Malnutrition (general)	Diminished host defenses against toxic effects.
11.2 Vitamin deficiency (selective)	Selective susceptibility (depending upon abnormality).
11.3 Inborn errors of metabolism	Certain genetic diseases associated with increased susceptibility to mutagenic effects.
12. Mental status	Stress may increase susceptibility to some toxic exposure. Stress, affective disorders, neuroses or psychoses may mask, mimic, or modify the clinical presentation.

---

The next step is to establish a relationship between the present illness and the pattern of work practices. Further inquiry may be necessary to document the association with work or the possibility of indirect exposure. Table 15.6 provides questions that may be used productively to follow up on positive responses or to evaluate other possibilities of exposure. The Essential Library for the Occupational and Environmental Physician in this book provides convenient sources of information for quick reference.

A common mistake in evaluating suspected occupational illness or injury is to focus too narrowly. A laceration from a minor accident may seem simple, but cases of trauma may not be as simple as they appear. Drowsiness or loss of coordination due to solvent toxicity or alcohol abuse may have played a role in the accident. A seemingly innocuous hyper-pigmented skin rash may be a cutaneous sign of arsenic toxicity; a complaint of shortness of breath may be pulmonary, cardiac, psychogenic, hematologic, toxic, musculoskeletal, or neurologic in origin.

Skin disorders are the most common occupational diseases. The OEM physician should pay close attention to the distribution of the lesions on the body, particularly with regard to clothing and exposure to sun or chemicals in the workplace. The entire body should be examined, not just the involved part, and the findings meticulously recorded with illustrations and, if possible, with photographs. Occupational skin disorders may change rapidly, making accurate documentation essential. Previous treatment should be specified in the notes. A reference book or atlas on occupational dermatology should be close at hand.

Eye problems are common and often accompany skin changes. Eye irritation often accompanies skin irritation from airborne or volatile irritants. Cataracts, corneal or lens discoloration, and neuro-ophthalmologic changes may indicate possible toxic exposures. Pupillary and oculomotor abnormalities suggest neurotoxicity or trauma until proven otherwise.

The physical examination should be thorough but directed toward the chief complaint. The stigma of alcohol or drug abuse may have a direct bearing on the case.

## **Referral**

As important as knowing the basics of occupational medicine is knowing when to refer a patient to a clinical specialist. Detective work is involved in evaluating obscure or complex cases. Such cases usually require the assistance of knowledgeable consultants. In general, referral

is indicated when a physician cannot confidently determine whether a problem is work-related, is unable or unwilling to prepare the necessary reports under workers' compensation, or is not prepared to manage an important aspect of a case knowledgeably. Important aspects of the case would include issues such as disability evaluation, antigen testing, assessing fitness to return to work, and screening for a particular problem such as asbestos-related conditions. The best course of action for a physician not systematically prepared in OEM is to complete the examination, record all findings in great detail, and then refer to a qualified consultant with a copy of the medical record and any pertinent information about exposure. There are few OEM specialists in private practice. In the absence of a reliable consultant in the community, it is usually most prudent to refer patients to OEM physicians affiliated with medical schools or large teaching hospitals, for reasons of quality assurance and accessibility to specialized tests.

OEM physicians are more involved, in general, in diagnosis than in treatment. Their expertise, while necessarily broad, cannot be detailed in every conceivable disorder that might have an occupational association. Occupational physicians cannot reasonably expect to be personally knowledgeable about all specialty areas. In cases in which one feels less certain about one's clinical acumen, it is therefore important to cooperate with a clinical specialist, preferably one knowledgeable enough to appreciate the occupational issues involved, and motivated enough to be thorough in answering the essential questions.

For the OEM specialist, on the other hand, referral to a clinical specialist and the dialogue that ensues may provide new insights into the particular problem, define more narrowly the differential diagnosis, assess impairment, generate new ideas about the case, and suggest further diagnostic procedures that are needed. Consultations with clinical specialists are often facilitated by outlining the critical questions explicitly in a concise letter of referral accompanied by reprints that address the toxicology or ergonomic issues involved. When frequent referrals are likely to arise in the future, it is desirable to cultivate a relationship with the clinical specialist by sending

along clippings, reprints, or course announcements dealing with occupational aspects of his or her specialty, and inviting the specialist to teach or participate in clinical conferences where occupational issues are addressed. It is particularly useful for most occupational physicians to have access to specialists they can trust in orthopedics, rehabilitation medicine, dermatology, pulmonary medicine, and neurology.

Every physician develops a personal style of performing the physical examination, and performance of this procedure becomes nearly automatic for patients without specific complaints.

### **Intervention in the Workplace**

Some basic steps to be taken by the physician in managing occupational health problems on behalf of a patient are outlined in Table 15.9. The worker should first be advised of his or her rights under the law. It is against the law for an employer to fire an employee for refusing to work in a hazardous situation. It is a right of employees as a group to petition the appropriate government regulatory agency for a workplace inspection. Once the worker is aware of these rights under the law, however, a certain degree of judgment is required to use those rights constructively. Despite guarantees of legal protection, workers may still find their jobs threatened. Employees who complain of unsafe work practices may be singled out for attention and fired, with other reasons given as an excuse.

An open line of communication with the employer is in the best interest of the worker, and is often essential to resolution of the problem. Except for adherence to legal reporting requirements, the employer should generally be given an opportunity to correct the problem before any regulatory agencies become involved. Attempts to “blind-side” the employer by withholding information or by proceeding immediately to filing legal claims often provokes an uncooperative and adversarial posture. Proceeding methodically and assuming that the employer is committed to resolving the

**Table 15.9.** Checklist of Occupational Health Interventions by the Physician Acting on Behalf of the Patient

*Preliminary Steps*

1. Report problems to the appropriate agency.
2. Evaluate urgency of the problem.
  - 2.1 Is condition serious to the patient?
  - 2.2 Are others in the workplace affected?
  - 2.3 Is there a pattern to the problem?
  - 2.4 Can the problem be explained by the characteristics of the working population or illness in the community?
3. Consider relationships to all parties; avoid conflicts of interest and observe duty to worker/patient.

*Intermediate Steps*

1. Prepare a complete report in writing, with as much information as is available.
  - 1.1 Consultation report to referring physician, if applicable.
  - 1.2 Update periodically with new notes as new information becomes available (do not rewrite unless initial report proven to be grossly inaccurate).
  - 1.3 Keep track of all versions, correspondence, and documentation (for legal and medical reasons).
  - 1.4 Avoid personal or judgmental comments.
  - 1.5 Identify essential issues explicitly.
  - 1.6 Identify further information or workup needed.
  - 1.7 Be meticulous with information that pertains to causation.
2. Review the literature in detail or refer the case to a qualified consultant.
3. Establish clear lines of communication with employer.
  - 3.1 Protect the worker.
  - 3.2 Avoid confrontation.
  - 3.3 Be helpful; indicate resources that can help resolve the problem.

problem is usually the best course, because it allows even a reluctant employer a face-saving way out of the dilemma and establishes the good faith effort on the part of the worker.

The single most important contribution of the physician in these cases is to provide complete and accurate documentation. Suspected cases of occupational injury or illness are reportable by law virtually everywhere in North America. This is usually done by completing the "Physician's First Report" form with the workers' compensation board or insurance carrier. Recording the facts as presented and observed in a more detailed case summary is an essential first step, particularly in cases that are likely to be disputed. At every subsequent step of the case, the details should be carefully recorded in the medical record. These records are subject to subpoena. Personal statements and extraneous information have no place in this or any other medical record.

The physician's primary role is to define the problem and to describe the health outcomes observed. A clinician without special training is usually in a poor position to prescribe specific controls. That function is best left to experts in safety and occupational (industrial) hygiene.

In identifying occupation-related diseases, it is important to look for patterns in the distribution of disease among workers in a particular workplace. Underlying demographic features of the workforce—age, sex, race, ethnicity, smoking habits, social class, alcohol and drug use patterns, and family relationships—may have a very strong effect on the relative risk for various health outcomes. An individual clinician or a group of practitioners in a given area may notice an unusually large number of cases of an uncommon disorder among workers in a specific industry or workplace. This is how the association between hepatic angiosarcoma and vinyl chloride came to be recognized. Such patterns may suggest an occupational association, but clusters of cases are not, in themselves, proof of an occupational hazard. Apparent clusters of disorders without identifiable causes are common, and are often the result of chance alone. A formal

epidemiologic study is usually needed to clarify the situation, and even this may not definitely rule out a chance association, especially if the disease is not rare.

If the problem is significant and satisfactory action is not forthcoming on the part of the employer, the worker along with other employees or through his or her union may request an inspection by the appropriate state, provincial, or federal regulatory agency. The names of workers initiating a request are kept confidential by the agency. Inspections by the U.S. Occupational Safety and Health Administration (and most state and Canadian provincial occupational health agencies) fall into four priority categories: (1) imminent danger, where a condition exists that might lead to death or serious physical harm; (2) fatality and catastrophe investigations; (3) investigation of complaints; and (4) routinely programmed inspections. Lower-priority inspections are handled on the basis of the severity of the complaint, and on the time and resources available. A complete inspection of the workplace is not always made in response to a complaint, especially when the workplace has been inspected recently on a routine basis. In such cases, partial inspections are focused on the source of the complaint.

An important aspect of managing occupational health problems is the psychological support of the patient. Patients who have experienced an occupational disorder often feel hostility, rage, and bitterness; these feelings can be magnified and augmented by the complexities and frustrations of the compensation system for dealing with often large and bureaucratic organizations. Additionally, these frustrations and problems are often compounded by a reluctance on the part of families and neighbors to accept the validity of the patient/worker's claims of illness, particularly if the impairment is not obvious. The result, all too often, is an embittered patient who feels wronged by society as well as by the employer. Psychological and emotional de-compensation is common. At least some of this can be helped by psychological support early in the process, and by counseling a worker with a disputed claim on the realistic odds of success as well as the limitations of the process.

Medical practice has been substantially changed by the introduction and widespread use of “practice guidelines.” Practice guidelines are recommendations for diagnosis, treatment, and management during recovery in uncomplicated cases or cases where the complications are predictable. They are specific for particular injuries or illnesses. They may be provided in the form of an algorithm, a narrative, a care map, a computerized decision support tool, or any other appropriate format. Most practice guidelines are developed by medical societies or authoritative organizations; however, they vary considerably in quality and detail.

The standard for the field of OEM is the *Occupational Medicine Practice Guideline* series developed by the American College of Occupational and Environmental Medicine, which is currently available in an updated second edition. (The series is called the *ACOEM Practice Guidelines*, or APG for short.) The *Guidelines* have been adopted in several states as “presumptively correct,” meaning that in any dispute regarding provision of care, the *Guidelines* are considered to be correct unless there is a good reason why they should not apply. A second, compatible set of benchmarks known as the Utilization Management Database is available for utilization review.

Practice guidelines are used by the clinician to get the best results in most cases. They demonstrably improve the practice of medicine by guiding treatment toward what works most reliably and is associated with the best outcomes, such as the highest functional recovery, most rapid recovery, and least residual disability. As such, they are designed for patients who fall within a certain range or scope, and who do not have co-morbid conditions that would interfere with the standard treatment plan. A good set of guidelines takes into account variation among patients and respects that complications may interfere with the management plan.

Practice guidelines are used by insurance carriers, including workers’ compensation carriers, as a guide for pre-approval and for utilization review. Sometimes they are applied too rigidly or without appreciation for individual patient variation. This is an abuse of the guidelines if the review does not take into account factors that make

a difference in care and outcome for an individual patient. However, it is not an abuse if physicians are given an opportunity to explain and justify why deviations from the guidelines make sense and are necessary for superior patient care. One advantage to having guidelines openly available is that physicians and insurance carriers have access to the same recommendations, and the process of pre-approval becomes more transparent.

From the legal point of view, practice guidelines are valuable (when their use is widespread) in establishing community standards of care. If guidelines are consistently followed when appropriate, and deviations are documented and justified, the physician is in a much stronger position to defend a decision.

Practice guidelines are the product of the movement toward evidence-based medicine that began in the UK, Canada, and the United States in the last third of the twentieth century. This movement had several roots, two of which are small-area analysis (which showed that one town or area might have markedly different rates of medical procedures with the same or poorer outcomes) and critical appraisal (essentially, a systematic way of accurately reading journal articles). Other significant influences were clinical epidemiology, evaluation of medical technology (applying cost-benefit analysis, although effective regulation never took hold), the rise of randomized clinical trials (as the gold standard for treatment), and the introduction of other statistical techniques, such as meta-analysis, in an effort to approximate the apparently definitive results of trials.

Over a period of three or four decades, a small group of academics, led by Dr. David Sackett of McMaster University (Ontario), advocated and operationalized the systematic review of medical knowledge. They pulled together the philosophy of “evidence-based medicine,” refined the approach by which the scientific literature is evaluated, and created the motivation to do it in practice through teaching and advocacy. They disseminated their views in scientific journals, but also in publications for the practitioner, such as the American College of Physicians’ *ACP Journal Club*, demonstrating the application of their methods as well as the ways in which particular findings fit in

context with the broader scientific literature. This led to efforts first to achieve consensus about best practices, and then, with increasing evidence, to weigh evidence of the superiority of one treatment over another. The movement reached its culmination in the Cochrane Collaboration, a network of international experts who do this work on a voluntary basis.

The evidence-based medicine movement was soon picked up by managed care organizations and insurance carriers, where it became a powerful business tool for achieving consistent results, achieving the best results with available resources, and cost containment. It quickly became evident that the best results for the patient usually resulted in the lowest cost to the system as a whole, at least within the scope of common medical problems. By standardizing care, managed care organizations were also able to manage their physician services and achieve flexibility in staffing. (Many feel that this was at the expense of physicians' professionalism and autonomy.)

Problems soon arose. Guidelines have attracted opposition in many quarters of medicine. They are perceived to constrain the physician as well as the practice, to suppress physician innovation, to ignore personal experience, and to be applied too rigidly in practice. They are widely viewed as adding to paperwork, although paperwork and the physician's liability are greater when outcomes are poorer. Proponents respond that medical innovations should be evaluated rigorously under experimental protocols. It is reasonable for physicians to vary in their practice within limits, taking into account the community being served and the patient's individual and family needs; however, the best treatment for the best outcome should always be provided unless there is a defensible reason to do otherwise. Any one physician's practice is unlikely to match the diversity of the world experience, as reflected in the medical literature, and is subject to bias from the referral source.

Implementation of guidelines has often been a battle. Even good guidelines got a reputation as a shield behind which crafty payers could hide, and as a bureaucratic nightmare. Utilization review and pre-authorization is sometimes inappropriate, and some reviewers

lack the medical knowledge to know why. Denials of claims for reimbursement, delays, and arbitrary decisions are sometimes blamed on guidelines, regardless of the real cause. Guidelines intended for simple cases are often inappropriately applied to complicated cases that they were not intended to address, in defiance of the language in the guideline. Abusive practices still occur and hurt the cause of sensible guidelines. When they are so bad as to discredit a good, solid guideline, even more harm is done because quality care for future patients is then compromised. However, misapplication is not a failure of guidelines; it is a failure of the system applying them. These issues have emerged with all guidelines, and the ACOEM *Practice Guidelines* have been no exception. To date, however, they have been accepted in practice with remarkably little resistance.

In 2000, David Sackett himself denounced the process (in an article in the *British Medical Journal*), stating that he was withdrawing from evidence-based medicine because the process had gone too far, and the opinion of experts carried too much weight in medical practice. For a while it looked as though the pendulum had swung as far as it would go, at least in general healthcare.

In the meantime, workers' compensation in most states was in crisis and needed some good ideas. As part of a general agenda to introduce managed care into workers' compensation medical coverage, practice guidelines became part of a reform movement, particularly in California.

There are unique evidentiary problems in OEM. Data on treatments and disease outcomes are fragmented and proprietary, held by the individual insurance companies (except in six states and all Canadian provinces), and are never pooled. There are very few randomized clinical trials, and it is almost impossible to implement a trial in a contemporary workplace setting. Good outcome studies are rare; such studies, when they are available, are often weak, uninformative, and incomplete, especially on secondary issues.

Occupational medicine guidelines cannot be uncritically adapted from their general medicine counterparts. Many of the common occupational conditions are shared with general medicine and surgery,

but the studies available in the general medical and surgical literature almost never address work-related issues. Such issues include standards for fitness to work, expected time to return to work, prognosis for second injury, accommodations that might be necessary during the first weeks back at work, and the risk of functional impairment. Sometimes the evidence is simply not there to support anyone's guidelines, in which case judicious use of consensus recommendations and other "soft" evidence becomes unavoidable.

The philosophy is that, properly constituted, good guidelines achieve better and more predictable results for the majority of patients, establish a standard of care, and place medicine on a defensible footing (cost-wise and in terms of liability). Properly implemented, they require the physician to justify and therefore to think through and document why something different should be done, allowing the physician to do so if there is a good reason. Properly framed, they extend only to a certain point and recognize that complications, multiple trauma, second injuries, co-morbidities, and unusual patient characteristics make the individual physician the best judge of how to proceed in exceptional cases.

Guidelines have grown in popularity and have proven their worth. What is now required is an integrated system centered on the injured worker that puts all parties on the same page: payers, providers, patients, utilization reviewers, and insurance commission regulators. Ideally, there will be an end to confusion and harmonization of the system, so that from the time care is authorized until the final bill is submitted everyone concerned, including the educated patient, will use the same, consistently updated database and the same rules.

## **PSYCHOLOGICAL ASPECTS OF OCCUPATIONAL DISORDERS**

Occupational disorders are always accompanied by some degree of psychological distress. In straightforward occupational injuries that can be expected to heal promptly, the psychological aspects may be minimal, but for more severe injuries and for all disabling injuries and

occupational illnesses, the presence of a psychological reaction to the condition is a constant in every patient. This psychological response varies greatly from individual to individual and is largely conditioned by personality factors, individual and collective beliefs, the exposure situation, and the perceived responsiveness of the employer. After certain devastating injuries (including those devastating to self-image and lifestyle), many injured workers show the characteristic symptoms of post-traumatic stress syndrome.

The psychological aspects of occupational disorders cannot be separated from the characteristics of the workers' compensation system (see Chapter 21). The features of the workers' compensation system, and similar systems in private insurance and entitlement programs, strongly shape the psychological reaction of the worker in chronic disability cases.

It must be emphasized that the presence of a strong psychological component in a case of suspected occupational disorder is not adequate reason to minimize the contribution of occupational factors in the case. Because a person is distraught and manifests anger toward an employer does not mean that the disorder is fictitious. Because a person is cool and controlled during interview does not mean that he or she is malingering. Because a person exaggerates the degree of pain or functional impairment that he or she experiences does not necessarily indicate a compensation neurosis or an attempt to mislead. Rather, it is exceedingly unusual to observe a significant case of suspected occupational disorder in which no psychological component is apparent. When these rare cases appear, denial masks the emotion underneath.

### **Psychological Dynamics**

The psychological dynamics of occupational disorders are often discussed but have been the subject of little critical or systematic research. Impressions are therefore based on repeated observation of individual patients, many of whom follow a very stereotyped pattern of behavior and emotional response following the injury. The degree

of the emotional response does not seem to correlate closely with the nature or degree of the injury. Some workers show great anger or depression following seemingly minor injuries while others seem to accept devastating illness or disability with seeming equanimity. On probing, however, similar dynamics are usually present among cases, but the individual reactions are very different.

For the sake of convenience, "injury" will be used to mean "harm" in a generic sense, and will be used in the following section to include both traumatic injuries and the onset of apparent illness. Assuming that a worker was in good psychological health before the injury, the injury is a discreet event that results in alienating the worker from his or her peers, the employer, and at the extreme, from family and community. At the very least, it involves an abrupt reassignment of family and social responsibilities as well as anxiety over the prognosis and implications of disability, particularly for financial support.

It is natural for people to want to assign responsibility for the injury. Legally, the identification of the responsible party is the basis of tort law (which the workers' compensation system was designed to replace; see Chapter 21). The most obvious party to blame for the injury is the employer, a supervisor, or a co-worker. This is perfectly natural and it is often justified, especially if there was some demonstrable fault involved. However, blaming is usually more destructive to the injured worker than to the party responsible for the injury. Blaming rarely results in correction of the problem or expedient handling of the claim, and it usually impedes both because it provokes a defensive reaction. An exaggerated tendency to blame, beyond the point of assigning sufficient responsibility to establish the relationship of the injury to the work, is often associated with very poor labor-management relations (and poor working conditions), a psychological profile of hostility and anger antedating the injury, or a dependency relationship with the employer (and usually others in the injured worker's life).

Perhaps less natural, and in some ways more psychologically damaging, is blaming oneself. Blaming oneself can seem to be a mature

response from an individual coming to grips with the injury; it has the outward appearance of taking responsibility for one's own actions. However, blaming oneself is a particularly sophisticated, inward form of "blaming the victim"; it is truly valid only when the worker consciously and intentionally committed an act that resulted in self-injury. Blaming oneself is not reasonable if the injury resulted from an action or moment of distraction that was not in itself negligent. This psychological reaction is often seen in people who have a tendency toward depression, or who have obsessive personality characteristics or a strong sense of personal responsibility. Self-blame often results in a loss of self-esteem and depression. It is generally far healthier to share the responsibility for the injury, especially when circumstances could have contributed to it, and to concentrate on taking responsibility for recovery and rehabilitation.

The injured worker usually assumes a sick role without the presence of a chronic illness or sustained decline in health. The typical injured worker goes abruptly from being an independent, self-directed wage earner (and, often, the support of another person or a family) to a dependent state (and at worst, an invalid), in which he or she becomes a passive recipient of what is often misunderstood as welfare. (Workers' compensation benefits are not welfare or social assistance, but rather are a form of insurance.) Most accept this state for a while, recover, and return to independence. Some do not, and either overreact to their new dependent state with anger or become passive and chronically dependent.

The short-term disability resulting from the injury immediately sets the worker apart from his or her peers. The worker is out of the social circle represented by work, and is now recovering at home or is otherwise engaged at a time when most adults in the community are at work. If work was a major focus in the life of the worker, this change results in an immediate separation and social handicap in addition to the physical impairment. This may be lessened at first if the worker socializes with friends from work and stays in the social network, but over time the bonds between the injured worker and fellow workers loosen and the social relationship becomes strained.

If the injury was severe, there may be anger and resentment on the part of the injured worker that fellow workers cannot share or maintain over time, if they are to do their own work. Fellow workers usually find it uncomfortable to maintain their relationship with an injured worker beyond a few weeks because they tire of stories about the injury, and because the workplace changes but the injured worker's mental image of it is frozen at the point in time when the injury put him or her off work. Additionally, resentment of supervisors or co-workers regarding the circumstances of the injury may be threatening to workers still employed there. If the injured worker is then neglected or shunned by co-workers, the sense of alienation may become overwhelming and the injured worker's resentment may become directed at them. This is one reason (among many) why early return to work is desirable if at all possible.

A serious injury is naturally followed by a sense of shock (during which the injured worker may cope remarkably well on "automatic pilot" despite unresolved conflict), followed by the familiar pattern of denial, rage, grief, and acceptance of the resulting disability. In occupational disorders, however, many factors interfere with the normal progression through these stages. In any disputed claim with a workers' compensation, there is a need to review the circumstances of the injury repeatedly while preparing for appeal or other litigation. Each time, the emotions return and are acted out, particularly in court testimony. The sense of rage must be maintained at some level in order to sustain the injured worker's motivation to persevere with the litigation. Because of the time required to process disputed claims, the injured worker may be off work but is constantly reminded of the injury and naturally seeks emotional relief through talking about it. Although some may listen with interest at first, their patience with the repeated story is soon exhausted and they turn away.

An occupational injury places great strains on the family. Routines are upset and the recovering injured worker may be underfoot at home. In traditional families, the male and female roles may be reversed as the injured husband stays home and the wife must work to support the family. This may be very threatening and may lead to

arguments, sexual problems, and self-doubts about masculinity. Sometimes there is acting-out behavior as the injured worker over-compensates or the spouse or other members of the family take advantage of the situation, sometimes in the context of serious pre-existing family problems.

The family is sometimes not capable of providing the unconditional support the injured worker expects to receive. Very quickly, family, friends, and community tire of the story and of the recriminations and blaming; they turn away, cease listening, or may even accuse the injured worker of laziness or of faking his or her impairment. The other members of the family, usually the spouse, may feel considerable anger at having to assume a greater share of support and having to deny personal ambitions; sometimes this spousal anger is acted out or expressed abusively, but it is almost always perceived by the injured worker. The injured worker, feeling more and more alienated, typically exaggerates the story further in order to get an empathic response, and directs rage at the party presumed responsible until the rage becomes bitterness. (This is the individual counterpart of the collective process described for “technical disasters” in Chapter 7.) Eventually, this may lead to a personality change in the injured worker as well as increased family tensions.

The cycle is further aggravated by the impersonal procedures of the workers’ compensation boards; these boards see themselves not as social service agencies but as business-like insurance companies. The workers’ compensation carrier or adjudicator may project an attitude of skepticism or disbelief toward a disputed claim; the worker interprets this as impugning his or her personal integrity. If the adjudication goes against the worker’s claims, this may be seen in the eyes of family and friends as evidence that the injured worker was attempting to abuse the system. Because disputed claims go to appeal, the injured worker must deal with impersonal quasi-legal formalities usually associated with determining guilt or innocence at trial. Although the procedure is intended to identify the work relationship of the disorder, it feels to the injured worker like a trial, especially if there is a challenge to his or her subjective complaints, factual narrative, or interpretation

of events. In order to carry on through this procedure, the injured worker has to sustain motivation either by anger, which often leads to inappropriate behavior, or desperation, which leads to pathological dependency on the process.

The phenomenon of "compensation neurosis" also complicates the picture. In anticipation of a disability award, the injured worker becomes preoccupied (and occasionally obsessed) with the degree of impairment. This may result in a paradoxical dependence on the impairment, since the injured worker's future well-being is associated with how severe the impairment may be. There are strong psychological pressures to remain fully impaired until the case is adjudicated; rehabilitation may even be perceived as threatening because it may reduce the projected amount of the future reward or the credibility of the claim. These motives and responses usually go on at a subconscious level and often interfere with counseling and other efforts to rehabilitate the injured worker. This is another reason why early settlement of the case is usually in the best interests of all concerned: it removes these subconscious barriers to recovery and allows the claim to be based on an objective evaluation of impairment before the compensation neurosis can complicate the assessment.

The psychological path of an injured worker need not be a downward spiral. One promising approach is the organization of self-help groups in which injured workers gather to talk about their concerns and feelings with the assistance of a mental health professional. Although there is some risk of mutual reinforcement of attitudes in this intervention, a properly led group may be able to come to insights in a sympathetic and supportive environment.

### **Subjective and Objective Findings**

Workers' compensation boards generally accept physical signs of injury at face value but are very reluctant to accept subjective symptoms as the basis for disability decisions. Some objective evidence of disability is usually necessary to substantiate a claim. There are several

reasons for this policy beyond the obvious one of not trusting the veracity of the claimant's report. One is that subjective findings cannot be quantified and therefore provide a very uncertain basis for making disability assessments, even in obvious cases. Another is that subjective perceptions of pain vary from one moment to the next and are amplified by attention on the injured part. Another is the fear that generous compensation awarded to persons with only subjective complaints may remove their motivation to work through their emotions and to get on with life and rehabilitation.

Exaggerated reports of pain and functional impairment are common in occupational medicine. The reasons some injured workers exaggerate their problem do not necessarily involve an attempt to deceive. The system is perceived by many injured workers as geared toward minimizing the degree of disability that will be accepted. (Proponents of the system may perceive it as striving to be objective, but this objectivity can sweep aside the injured worker's emotional response and seem so impersonal that the injured worker questions the motives behind it.) Often, there are also questions about whether the injury is in fact related to work. When a person untrained in law or science perceives that he or she is not believed or that the listener is skeptical, the usual response is to elaborate further and to become hyperbolic in description in an attempt to be more convincing. This may come across to the listener as obvious exaggeration and as further evidence that the speaker is trying to deceive. Something similar happens to the injured worker. Suspecting that his or her story will be dismissed, the injured worker will want to present as compelling a case as possible.

The most difficult and common discrepancy between subjective and objective findings in occupational medicine is the experience and description of pain. Pain is a complex subjective complaint, described in detail in Chapter 16. The nature of pain varies considerably depending on whether the injury is acute or chronic. Chronic pain, the most problematic and common subjective issue in the workers' compensation system, is mediated by different neurological pathways than acute pain. Acute pain serves a purpose in alerting the

organism to hazard and the need to tend to a fresh injury. Chronic pain has no known adaptive utility and is associated with depression and anxiety.

Chronic pain is subject to considerable amplification or minimization depending on the mood or focus of attention of the sufferer. When a chronic pain patient is made to laugh or is distracted, the sensation of pain diminishes, on a presumably neurological basis. If the attention of a chronic pain patient is drawn to the pain, such as by repeated questioning, the pain feels more intense and the prognosis more hopeless. This is one reason injured workers often feel worse at night in bed and have difficulty sleeping; in the absence of distracting daytime activities, their attention is focused on the pain. Likewise, when the patient is depressed, either by the circumstances of the injury or because of other factors in life, the pain seems more intense and intolerable. This apparent variability in the sensation of pain may lead to misunderstandings; it can be interpreted as evidence of malingering or exaggeration. However, it also leads to effective treatment of chronic pain through measures designed to alleviate depression and to encourage the injured worker to participate in activities.

Sometimes subjective factors act through muscle stiffness or the anticipation of pain to limit mobility. These obstacles to recovery can be removed by skillful physical therapy.

### **Somatization**

Somatization is the presentation of physical complaints that have little or no organic basis, but reflect psychological processes or disturbances. The complaints are real to the patient but have little or no relationship to objective findings or the state of health. As encountered in OEM, they are typically expressions of anxiety or depression.

Somatization is responsible for a high proportion of visits to physicians. Typical presentations include unexplained pain, unexplained nausea, palpitations, and other vague symptoms. Because of the nature of occupational injuries and illnesses, the typical range of

somatization symptoms includes pain, shortness of breath, dizziness, and weakness.

Somatization as a process should not be confused with hypochondriasis or somatization disorder. Hypochondriasis is a fixed belief that one is ill despite reassurance. Somatization disorder is a psychiatric diagnosis (DSM-IV 300.81) characterized by at least thirteen reported symptoms unresponsive to treatment, typically diagnosed in women and associated with personality disorders as well as with suicide attempts. (Because of the prevalence of somatization disorder among women and the historical connection between this diagnosis and the old rubrics of "hysteria" and de Briquet's syndrome, the diagnostic label of "somatization disorder" has been denounced as sexist. However, many psychiatrists believe that the evidence for a distinct entity is overwhelming, and that the diagnosis has its male counterpart in impulse control disorders.)

Somatization as a phenomenon in workers' compensation seems to be more fluid. Some generalizations have to be made on the basis of experience because systematic studies have not been done. Somatization in this context is diagnosed in approximately the same number of men as women, and appears to be more common in individuals who hold fixed beliefs about values and personal behavior, and whose moral standards are absolutist and possibly unattainable. Often, these individuals are socially constrained in an environment such as the military, police, or medicine that does not allow much individuality of action or open expression of emotion. It would therefore appear that this is a displacement, a safe way for people with rigid personalities or in rigid environments to channel a response to stress. This leads to a paradox. People who are least likely to accept the validity of emotional reactions and who perceive responses to stress as weakness will not accept that their condition is somatization. This makes them very difficult to treat and help. One way to reach them is to suggest psychological assistance for the stress that their condition is causing them, rather than for intervention in the condition itself. This opens the door for a qualified mental health professional to work with them.

The injured worker, or the worker who is fearful of a toxic exposure, may become fixated on his or her health. Those workers who do so may become exquisitely aware of small changes in their bodies and of passing sensations that other people ignore in the distractions of daily life. These passing changes may include modest swelling of the hands and feet, vasomotor changes in the skin, stiff muscles, goosebumps, transient itches, momentary lapses of memory, muscle twitches, and all of the other short-lived phenomena in the human body of which a person is normally unaware.

A common manifestation of somatization in occupational and environmental medicine is the rapid appearance of symptoms (that are genuinely distressing to the patient) upon entering the workplace under circumstances that are implausible for toxic or immunologic mechanisms. For example, a worker may experience the immediate onset (or onset within one-half hour) of headache and lightheadedness upon reentering a workplace in which solvent concentrations are not nearly sufficient to produce the effect. Another may become short of breath in anticipation of entering a room. (In one memorable incident, a worker developed acute abdominal pain requiring admission to hospital while walking past the door of a laboratory in which he suspected that he had once experienced toxicity.) In each case, the subjective symptoms are implausibly related to exposure and do not relate to objective indications of disease. These cases usually have a component of acute or strongly suspected toxicity in the past; usually, the toxicity has affected the worker in the same workplace, and the responses are reported to be consistent and (in the mind of the worker) reproducible. In these cases it appears that there may be some degree of learned (or behaviorally conditioned) behavior in which the stimulus of the workplace environment results in a conditioned response that takes somatic form. The symptoms appear to be distressing to the worker involved and may be similar in mechanism to phobias and possibly panic disorder, mediated at least in part by the sympathetic nervous system discharge in response to stress.

Essential in such cases is a convincing screening medical evaluation. This is important both to rule out true organic disease and, later, to

convince the patient that there is indeed no organic basis for the complaint. The medical literature is full of cases in which subtle findings have led to the detection or foreshadowing of a serious disease; patients are well aware of this and readily believe that they will be the unique case among thousands. The screening medical evaluation should therefore be sound, thorough, and convincing. It is axiomatic in clinical medicine that even hypochondriacs get sick, and when they do, they are at increased risk because nobody takes them seriously. The same is true with somatization.

Management of the somatization component of an occupational disorder can be quite challenging. Because of the complex psychological processes that inevitably accompany occupational injuries, denial by the physician of the validity of the symptoms is usually interpreted by the worker as denial of the validity of all aspects of his or her case and an impeachment of integrity. One approach that seems to work in this situation is for the physician to emphasize that the symptoms are real; it is the mechanism that produces them that is being questioned. A physician can be sympathetic to the distress of the patient without agreeing that the symptoms are the result of an exotic pathological process or a toxic exposure. One way to do this is to propose a legitimate alternative mechanism as the cause of the symptoms. The somatization symptoms should not be ignored when the patient complains, but the physician should not dwell on them. Instead, frequent expressions of reassurance and reinforcement of improvement may help. Treatment of underlying depression and anxiety is also important, with frequent visits or telephone contacts for reassurance.

### **Conversion Reactions**

Unlike somatization, conversion reactions present objective signs that have no basis in organic disease. They are not misperceptions, as in the case of somatized symptoms, but rather present as outwardly visible dysfunction, such as loss of ability to use a limb or torticollis. The person is not conscious of the nature of the action and truly

believes that the impairment results from an injury. When distracted, the person usually shows normal function of the affected part, which lacks atrophy or other signs of chronic immobility. Conversion reactions are unusual in occupational medicine. They usually follow a stressful event in the life of the individual, and are thought to represent socially acceptable ways of dealing with otherwise intolerable conflicts. Patients with conversion reactions may also have somatization, but the two should not be confused. Somatization is subjective in nature. The prognosis is highly variable, with some cases leading to serious depression. Again, the nature of the disorder makes it hard to study, and few studies have been attempted in OEM.

### **Malingering**

Overt malingering, in the form of entirely falsified claims based on nonexistent injuries or medical conditions, is probably rare in OEM overall but is seen disproportionately in certain settings. Much more common are attempts to misrepresent the severity of existing disorders to inflate the degree of impairment, or to claim that an injury was work-related when in fact it was not. These attempts at partial malingering should be distinguished from the sincere, if exaggerated, description of symptoms and from somatization, neither of which is a conscious attempt to deceive.

Some individuals deliberately attempt to mislead or to inappropriately exaggerate their conditions; the psychological background to malingering is also more complex than is usually recognized. Some of these people may have sociopathic personality disorders that allow them to be particularly convincing. Others, however, may have a dependent personality disorder that expresses itself as an internalized belief that the world should provide support and comfort without one having to assume adult responsibilities. Some malingering individuals have antisocial personalities and simply attempt to abuse the system.

A particularly troublesome group of malingerers consists of those who do not see the wrong in the filing of a false claim, because the social norms in their communities lead them to feel entitled to a

pension or welfare support. These individuals may see themselves as upright and honest and only claiming that to which they feel entitled. This is particularly likely when these people live in a community where many residents are already receiving workers' compensation benefits. These applicants may feel no compunction about misrepresenting their cases because they feel that their lack of a compensable injury should not disqualify them; they are as deserving morally as their neighbors, many of whom they may believe lack meritorious claims. Workers' compensation (which is actually a form of insurance) is more popular than welfare benefits because benefits are greater.

Malingering should be suspected when the objective findings do not match subjective complaints in a disorder in which this would be expected, when the patient shows signs of antisocial or sociopathic personality disorder, or when the patient's approach is inappropriately casual or flippant. Other signs of malingering may be when the patient is uncooperative in an examination or cooperative only until an objective test is proposed, when the symptoms reported are inconsistent and the inconsistency is explained with elaborate and defensive explanations, and when the complaint of disability seems out of proportion to the injury supposedly sustained. None of these characteristics establishes malingering, but a combination is suggestive. There are many "tricks of the trade" used to identify malingerers through clinical testing; these depend on observing the subject performing an action that would not be possible if the complaint were authentic. However, these tests cannot prove how an injury originated, as in the case of a malingerer who falsely claims that a genuine injury was work-related.

There is a unique risk with such patients, and that is that they sometimes claim that the physician injured them in the course of conducting an examination. The OEM physician faced with a patient who seems to be anticipating a misadventure or who is complaining of increased pain because of some maneuver during the examination should take steps to protect him- or herself. This may mean bringing an observer into the room, perhaps on the pretext of showing them some finding, documenting everything that is done in meticulous

fashion, proceeding slowly with the examination, and avoiding any maneuver that is likely to cause pain but documenting the reason it was not performed.

## **MULTIPLE CHEMICAL SENSITIVITY AND RELATED BELIEF SYSTEMS**

“Multiple chemical sensitivity” (MCS) is a highly controversial diagnosis. Most informed mainstream medical practitioners do not believe that it exists as a physical disorder or a legitimate medical diagnosis. Some do accept it as a pattern of behavior without accepting that it has a physical basis. Recognition of MCS as a distinct disease, however, has been advocated by a group of physicians who initially called themselves “clinical ecologists” (see Chapter 1).

MCS consists of a pattern of intolerance to different chemicals, not respecting chemical class or toxicity, that manifests itself in a number of largely subjective symptoms. There is no generally accepted case definition. Clinical ecologists (who now prefer to be called physicians practicing “environmental medicine”) believe that there is an organic basis for the disorder, variously attributing it to damage to the immune system, neural connections between olfaction and the limbic system, degenerative neurological conditions, nutritional imbalance, and food allergies.

Clinical ecology, to use the original name, is an unrecognized medical specialty, operating outside the usual specialty boards and structures with a network of interlocking boards and foundations. It has attracted adherents in large numbers since about 1962, when a textbook was published. The American Academy of Environmental Medicine, founded in 1965, is the chief organization of practitioners who diagnose MCS and its progeny. It has received great attention in the mass media. The fundamental tenet of clinical ecology is that scientifically documented patterns of allergy and immunological response are too narrowly perceived, and that it is possible for the human body to be exquisitely sensitive to minute exposures to chemicals not previously recognized to be toxic or antigenic. In

practice, clinical ecologists diagnose a wide variety of conditions as immunological in nature, in response to exposures to numerous chemicals encountered in the environment (especially pesticides) and in food. Chemical exposures are not only antigenic, in the interpretation of clinical ecology, but they may also be immunotoxic, modifying and depressing the immune response and rendering the persons so afflicted more susceptible to infections (such as yeast infections) and to aberrant reactions to other chemicals.

An increasing number of claims related to chemical exposures in the workplace have been received by workers' compensation boards; typically, these complaints concern the induction of more general sensitivity to environmental chemicals following an exposure incident or a prolonged exposure.

The patient population of clinical ecologists tends to consist of younger, health-conscious adults. Children are rarely treated. The standard methods of diagnosis in clinical ecology are empirical, using provocation or withdrawal of foods to identify putative allergens. Methods of treatment vary but may include neutralization therapy, chelation, nutritional counseling, "detoxification," and a variety of unorthodox, nominally "holistic" treatments. (When challenged on the efficacy of these treatments, most practitioners attempt to cite evidence, but others have claimed that efficacy is unimportant because they were making creative use of a placebo effect.)

Over time, the terminology has changed and continues to change. MCS was originally called "environmental hypersensitivity disorder," "multiple chemical sensitivities," "ecologic illness," and, most dramatically, "20th Century disease" (the term was abandoned in time to keep it from becoming out of date). The movement to legitimize the diagnosis seems to have advanced to other, equally questionable diagnoses, most recently "mold illness," "toxic-induced loss of tolerance" (TILT), "toxic encephalopathy," and trace element imbalance. In recent years the number of patients claiming to have this disorder seems to have declined as the number of cases allegedly due to mycotoxins from mold exposure increased. "Toxic mold" cases appear to

be on the decline, with “toxic encephalopathy” becoming the more popular diagnosis at the time of this writing (2009).

MCS is said to be a chronic disorder characterized by an intolerance to many chemicals normally encountered in modern life at levels not previously reported or suspected to cause health effects or discomfort in most people. Particular emphasis is placed on exposures in combination and interactive effects. The adverse effects range widely but are subjective in nature, are usually thought to affect the central nervous system, and are not associated with objective reproducible findings or abnormal findings on conventional clinical laboratory tests. The literature of clinical ecology is inconsistent in demonstrating clinical features that might be specific in making this diagnosis. At the extreme, patients with this diagnosis may withdraw from human society, live in amenity-impoverished rooms or even plastic bubbles, and restrict their diet to a very narrow range of monotonous foods. Their lives become essentially deprived and focused on avoiding chemicals. In the few studies that have examined the outcomes of such situations, the patients do not appear to improve with the treatment given and may even report a proliferation of symptoms.

It has been difficult to evaluate the teachings of clinical ecology because the case definitions are inconsistent, the rationale is constantly changing, and the vocabulary does not respect standard usage. (“Toxic encephalopathy,” for example, is assumed in this definition to occur in patients who have no obvious neurological impairment—although in mainstream medicine, the term “encephalopathy” implies a devastating, diffuse brain disorder with obvious signs.)

On the other hand, there are well-known mechanisms of sensitization and response, including asthma and allergic reactions, and irritation effects when these mechanisms are superimposed on reactive airways or inflamed mucosa. Theories of clinical ecology incorporate these common disorders indiscriminately, combining asthma and atopy with MCS and ill-defined conditions such as Gulf War syndrome in a theoretical construct that has little coherency.

There is no question that there are many mechanisms of human illness and discomfort involving chemicals beyond the obvious allergic responses and recognized toxicological mechanisms. These include nonspecific irritation and inflammation, complicated presentations of airway reactivity, reflex-mediated vagal reactions to odors (resulting in nausea), posttraumatic stress disorder and consequent depression and distraction, and conditioned responses (discussed in this chapter in the section on somatization). There is also no question that there is a distribution of susceptibility within the population, both among individuals with allergies (illustrated by the wide range of severity among asthma patients reactive to the same antigen) and individuals exposed to irritants (illustrated by the wide range of responses of nonsensitized individuals exposed to formaldehyde). Neurotoxic exposures do exist, and some cloud consciousness. Additionally, there is increasing evidence in the mainstream, accepted scientific literature that some chemicals are associated with modification of the immune response. All of these mechanisms exist, but no coherent theory of interaction has yet been formulated that convincingly explains why very low exposure levels of chemicals that are not only common in the modern environment but have been for many years should provoke such a bizarre and inconsistent response in adulthood.

Clinical studies of patients carrying the diagnosis of environmental hypersensitivity strongly suggest a consistent psychological profile. Depression and high levels of anxiety have been observed, as well as high scores on test scales designed to measure both hypochondriasis and somatization (but which also reflect health consciousness). A large proportion of these patients have histories of psychiatric illness or life trauma compared to the general population. Because of the dramatic withdrawal from normal social intercourse that occurs in many of these patients, and because of the vague, somewhat central-appearing symptoms, the condition bears obvious similarities to agoraphobia and panic disorder. A major problem with studies to date, however, is that they have not allowed the opportunity to follow a group of these patients longitudinally, either to determine who in

a cohort is at risk or what the natural history of the condition might be. Consequently, it is not presently possible to say with assurance whether the psychological features of the disorder preceded the clinical condition or are a result of it.

If, as seems to be the case, many of these patients are psychologically vulnerable, they are also easily victimized. Recruiting them into a self-referential world in which only the practitioner understands them and where the explanation for their condition lies in esoteric knowledge that keeps them apart from friends, family, and mainstream medicine is unlikely to help their condition. There is a risk of learned dependency in such situations that may deepen the psychological risk and isolation that these people must feel. This is a group of patients who need caring and compassionate medical evaluation and support. It is tempting to interpret environmental hypersensitivity as solely a problem of "chemophobic agoraphobia," manifested as a socially acceptable form of somatization in a population of anxious adults. This may well be true for some, if not the majority of MCS patients. Certainly the medical community interested in clinical ecology has not convincingly demonstrated the scientific basis for its system of thought or even the existence of the condition. However, there may be mechanisms of pathophysiology or subjective states of awareness that are unrecognized or underappreciated by mainstream clinical medicine. The conscientious OEM physician must therefore keep an open mind while remaining skeptical and insisting on evidence for novel disease theories. It is a difficult balance.

## THE ACUTE CARE CYCLE

Clinical counters in occupational medicine follow a predictable cycle, outlined in Table 15.10, for the individual injured worker. The cycle begins with the baseline state of health, proceeds through clinical management and rehabilitation, and ends with fitness to return to work or permanent disability—but in either case, to a new baseline.

Entering the cycle, the worker is in his or her previous state of health when experiencing an injury or illness. Prior to the onset of

**Table 15.10.** The Acute Care Cycle

- 
- 0.0 Worker in usual state of health until injury occurs or disorder is symptomatic
  - 1.0 Provision of immediate care
  - 1.1 Access and triage
  - 1.2 Quality and consistency of care
  - 1.3 Care maps and clinical algorithms
  - 1.4 Expectations, goals and objectives
  - 1.5 Entry into a care management system (ideal)
  - 2.0 Provision of follow-up care
  - 2.1 Role of case manager
  - 2.2 Case maps and adjustment for complications, individual issues
  - 2.3 Short-term disability
  - 2.4 Rehabilitation services
  - 2.5 Monitoring functional status
  - 2.6 Coordination of services and healthcare
  - 3.0 Conclusion of care
  - 3.1 Fitness to return to work, impairment assessment
  - 3.2 Accommodation or re-training
  - 3.3 Permanent impairment, long-term disability
  - 3.4 Chronic case services
  - 4.0 Performance of system
  - 4.1 Outcomes
  - 4.2 Client satisfaction
  - 4.3 Quality control/quality assurance
  - 4.4 Cost
  - 4.5 Risk of re-injury
  - 4.6 Successful job performance after return to work, or job loss
- 

the disorder, the worker is not a “patient” in the sense of a person needing or receiving healthcare; they may have been a “client,” benefiting from services for occupational health protection and wellness, of course, but they were not a patient in the accepted sense.

A work-related injury or illness represents a failure of occupational health and safety, because all occupational disorders are preventable, so there is immediately an issue of accountability, but that is not

addressed at this point. Once the injury or illness occurs, the priority for the individual injured worker, now a patient in the healthcare system, is diagnosis and treatment of the acute disorder. This is followed by a decision with respect to the relationship between the injury or illness and work (which is usually obvious in the case of injury). Unless the association is obvious, this process requires expert evaluation in a process called “analysis of causation.” (See Chapters 21 and 23.) However, that comes later.

In the initial encounter to receive acute care, the first issue is access and triage. In some American states, the employer designates the physician for the initial encounter, and in other states the worker chooses. In Canada, it is uniformly the worker’s choice. In either case, the healthcare provider should be aware of the special needs of occupational healthcare (specifically, the emphasis on prevention of disability and early and safe return to work as part of recovery), the requirements of the workers’ compensation system, and the need for accurate documentation and expeditious management. Many practitioners are not, and in the United States many physicians refuse to take workers’ compensation cases out of a misguided fear of the paperwork required and of getting caught up in a time-consuming appeals process.

The care that results in the least permanent disability and the earliest return to work that is safe for the worker is the best quality of care. For this reason, it is essential to benchmark care against best practices and evidence. Quality and consistency of service remain a challenge in healthcare. Guidelines, clinical algorithms, care maps, and other tools of evidence-based medicine have become essential to the standard of practice for occupational medicine. The standard of practice in evidence-based occupational medicine, unequivocally, is the guidance of the *ACOEM Practice Guidelines (APG)*.

Critically important is communication, first to the injured worker and then to the employer, regarding what to expect in the way of recovery and time off. Injured workers, like all patients, need to have goals, a sense of proportion, and an understanding of the timeline they may expect for recovery in order to plan their lives. The OEM

physician should communicate a few key points clearly, and should repeat them at the time of the first encounter for acute care to be sure that the injured worker understands them. These key points include:

- A sense of how serious the injury or illness may be. Without this, some injured workers leave with the impression that their injury is more serious than it really is and become demoralized.
- Goals for recovery. If the most likely outcome is full recovery, the physician should say so immediately. A goal is neither a promise nor a contract. It is a target toward which the physician and the injured worker will aim, one that will make sense out of further treatment, rehabilitation, and eventually determination of fitness for duty.
- Expectations of the injured worker. The injured worker cannot be passive in his or her own care. Reasonable expectations for adherence in treatment, for monitoring his or her own condition, and for return to work should be articulated in a matter-of-fact manner, and questions should be asked of the injured worker to make sure that he or she understands.
- Expected time to recovery. It is not necessary to be exact at this stage in treatment, but the injured worker should always leave with a clear idea of whether recovery will take days, weeks, or months.
- Employer's role. Employers decide when an injured worker will return to work but they need guidance on when the employee is ready. The injured worker must understand that the physician needs to report the injured worker's progress not only to the insurance carrier or HMO but also to the employer, so that return to work will be smooth and unopposed.

This conversation should not be delayed until a follow-up visit. For the most fortunate injured workers, a little reassurance that an injury is minor and will heal quickly is enough. Injured workers with

an injury that takes time to heal, or with an illness that may have a long time course or that may become chronic, need motivation and encouragement. They will be more likely to adhere to treatment and monitoring guidelines if they understand the eventual goal. Expectations and goals are also helpful in getting all of the care providers, including surgeons and medical specialists, on the same page with respect to a treatment plan.

A preliminary estimate of how long the injured worker is likely to remain off work can be derived by using the *APG* or the *Medical Disability Advisor*. Certain states, such as Minnesota, have compiled statistics on average time off work for specific work-related injuries. When the time comes to return to work, it will not come as a surprise if the time frame is communicated ahead of time, and there will be less incentive for the common complication of asking for additional time off. The injured worker should also understand that reentry into the workplace must involve the employer, and this means that the employer should be kept informed of the degree of progress in treatment and rehabilitation. These key points need to be made at the time of the initial clinical encounter for acute care. If they are delayed, misunderstandings are inevitable.

Ideally, at the time of the first encounter or within a week of initial treatment, the injured worker would enter into a comprehensive care management system. Such systems involve a case manager, usually a nurse, who schedules follow-up and specialist appointments, monitors whether the injured worker is keeping the appointments, confirms adherence to treatment and rehabilitation, talks with the injured worker regularly, coordinates the opinions and treatment plans of various specialists, and decides when the injured worker is ready for the fitness-for-duty evaluation. This case manager also identifies and removes obstacles to receiving care; this may include arranging for or paying for transportation to and from medical and rehabilitation appointments. Unfortunately, such systems are not available outside of existing networks.

Monitoring progress is very difficult within the current fragmented healthcare system. Monitoring is essential in following the progress

of the injured worker in order to identify obstacles to care or continued treatment, to monitor adherence, to identify complications or poor response to treatment, and to determine when the injured worker is ready for a fitness-for-duty evaluation in preparation for return to work. Workers' compensation carriers sometimes take a very directive approach, using case managers to follow the progress of injured workers and to coordinate care, such as specialist appointments and rehabilitation visits. Others expect the medical care provider to do this, which is generally unrealistic given the way that healthcare is currently organized and compensated. Very often, this task is performed by the physical therapist, who sees the injured worker regularly and evaluates function at every visit. However, rehabilitation services are often capped, and after the allowable number of visits is exceeded the continued services of the physical therapist may not be available.

Follow-up care has a tendency to lose focus after a few weeks. It is essential that the "care map" or treatment plan be reviewed with the injured worker from time to time, with a clear indication of where he or she is on the timeline. This motivates adherence and shapes expectations with respect to return to work.

The ideal situation is one in which the injured worker can return to work with accommodation while treatment is proceeding; this is the best scenario because it facilitates reentry and prevents long-term disability. However, many employers do not permit this. Employers should receive periodic updates on the progress of the injured worker so that they can plan for his or her return. As time goes on, the expected day of recovery will become firmer and projections will be more exact.

Expectations and goals also involve the employer. The employer may have a policy favoring early and safe return to work with accommodation for temporary disability, or may allow only fully recovered employees to return to work. The physician may influence the decision by providing regular reports on progress, without revealing confidential information, and by advising the employer when the injured worker is coming to the end of treatment or rehabilitation and is almost ready to return to work.

As the injured worker finishes a successful treatment plan and course of rehabilitation, a fitness-for-duty evaluation is required. This evaluation provides a determination of work capacity (as discussed in Chapter 18). The recovering worker may be fit, unfit, or fit with accommodation. If the worker is fit with accommodation, the evaluation identifies what form of accommodation, such as a change in the workplace, a limit on lifting, or adjustment of hours, may allow the worker to return to his or her usual job. If the worker is unfit, which at this stage implies permanent impairment, the evaluation determines the level of impairment, in preparation for calculating disability under workers' compensation. The evaluation may also identify special needs for chronic care, if required.

Throughout the cycle, the performance of the system as a whole needs to be evaluated on the basis of outcomes, not just care delivered. Other evaluation criteria influence client satisfaction (on the part of the injured worker), employer satisfaction (with the process and timely reporting), quality control and assurance, cost, and the risk of re-injury. Of particular importance is the risk of poor job performance. If an injured worker after treatment and rehabilitation returns to work but cannot do the job, the care cycle has failed him or her, regardless of the medical outcome.

## RESOURCES

- Aronoff GM, Mandel S, Gevese E, Maitz EA, Dorts AJ, Klimek EH, Staats TE. Evaluating malingering in contested injury or illness. *Pain Pract.* 2007;7(2):94–102.
- McCunney RJ, ed. *A Practical Approach to Occupational and Environmental Medicine*. Philadelphia: Lippincott Williams & Wilkins; 2003.
- Occupational and Environmental Health Committee of the American Lung Association of San Diego and Imperial Counties. Taking the occupational and environmental history. *Ann Intern Med.* 1983;99:641–651. (Source for Figures 15.1 and 15.2.)
- Rogers R, ed. *Clinical Assessment of Malingering and Deception*. New York: The Guilford Press; 1988.
- Rosenstock L, Cullen M, Brodkin CA, Redlich C. *Textbook of Clinical Occupational and Environmental Medicine*. 2nd ed. Philadelphia: Saunders; 2004. [Excellent book and very expensive.]

## NOTEWORTHY READINGS

- Almada SJ, et al. Neuroticism and cynicism and risk of death in middle-aged men: the Western Electric Study. *Psychosomatic Med.* 1991;53:165-175.
- American College of Physicians. Occupational and environmental medicine: the internist's role. *Ann Intern Med.* 1990; 113:974-982.
- Brown CS, et al. A practical update on anxiety disorders and their pharmacologic treatment. *Arch Intern Med.* 1991;151:873-884.
- Buckelew SP, et al. Health locus of control, gender differences and adjustment to persistent pain. *Pain.* 1990;42:287-294.
- Chapman SL, Brena SF. Patterns of conscious failure to provide accurate self-report data in patients with low back pain. *Clin J Pain.* 1990;6:178-190.
- Cheatle MD, Brady JP, Ruland T. Chronic low back pain, depression, and attributional style. *Clin J Pain.* 1990;6:114-117.
- Crawshaw R. Thoughts about diagnosis. *J Clin Epidemiol.* 1990;43:723-725.
- Cullen MR, Cherniack MG, Rosenstock L. Occupational medicine (in two parts). *N Engl J Med.* 1990;322:594-601.
- Dualba L, Scott RL. Somatization and malingering for workers' compensation applicants: a cross-cultural MMPI study. *J Clin Psychol.* 1993;49(6):913-917.
- Katon W. Depression: somatization and social factors. *J Fam Pract.* 1988;27:579-580.
- Katon W, et al. Somatization: a spectrum of severity. *Am J Psychiatry.* 1991;148:34-40.
- Kennedy HG. Fatigue and fatigability. *Brit J Psychiatry.* 1988;153:1-5.
- Leavitt F. The role of psychological disturbance in extending disability time among compensable back injured industrial workers. *J Psychosom Res.* 1990;34:447-453.
- Margo KL, Margo GM. The problem of somatization in family practice. *Am Fam Phys.* 1994;49(8):1873-1879.
- Puente AE. Relationship between self-consciousness, autonomic perception, and psychosomatic disorders. *Percept Motor Skills.* 1984;59:769-770.
- Radcliffe MJ. Unexplained illness: the mind versus the environment. *J Roy Soc Med.* 1995;88:678-679.

# **16 OCCUPATIONAL INJURIES AND MUSCULOSKELETAL DISORDERS**

In general medicine, as in colloquial language, an injury is a traumatic event or strain and usually involves the musculoskeletal system. However, the definitions in common use in workers' compensation and federal regulatory policy distinguish between injury and musculoskeletal disorders, which are often called "musculoskeletal diseases" in this context. That is because the underlying theory of disease differs between the two. This difference derives from workers' compensation, where the distinction between an event and a prolonged process in which the outcome is only the end result is important in establishing relationship to work, apportioning responsibility, and setting policies for compensation.

An injury, in the context of workers' compensation, is the result of a discrete event, such as a fall from height, laceration, strain, or fracture. A "disease," in this context, is the result of an ongoing process. (Acute poisoning, although usually considered a disease, is considered an injury in workers' compensation.) In workers' compensation, repetitive strain injury (also called "cumulative trauma disorder") is counted as a disease, not as an injury. It is considered to be the outcome of the cumulative effect of numerous small injuries, an approach that allows apportionment and conforms to legal definitions of injury. The

principles of care for minor trauma are well known to most physicians and do not need to be repeated here. This chapter limits itself to an overview of occupational injuries and the issues that make management different from that for injuries arising in the community. The emergency, medical, surgical, and rehabilitation management techniques for acute injuries are therefore outside the scope of this discussion. Musculoskeletal disorders are less familiar but are central to the practice of occupational and environmental medicine (OEM) and so will be discussed in detail.

For guidance on clinical management, the reader is referred to the American College of Occupational and Environmental Medicine's *Occupational Medicine Practice Guidelines* (referred to as *ACOEM Practice Guidelines*, or *APG* for short). The *APG* provides guidance to physicians on occupational injuries:

- Diagnostic criteria
- Criteria for assessing causation
- Consequences of disability and functional impairment
- Best practices for initial management
- Therapeutic modalities
- Rehabilitation
- Complications and deviations from normal recovery
- Disability management
- Documentation on evidence
- Algorithms for many disorders.

A major difference in management between acute injury and occupational musculoskeletal disorders is that it is not necessary to have a specific, anatomic diagnosis for the effective initial management of musculoskeletal disorders. Management rests on the control of inflammation, pain, and factors that place the injured worker at risk of chronicity, not on specific interventions. Knowing the anatomy is critical in suturing, setting a fracture, or injecting a joint. However,

knowing whether low-back pain is a problem with the facet joint or the disc, and whether upper extremity pain is a sign of early carpal tunnel syndrome or an evolving tendonitis generally makes no difference in the first instance: the patient is treated the same in each case and the outcome is generally good.

### **OCCUPATIONAL INJURIES**

Traumatic musculoskeletal injuries are the most common type of occupational disorder. Acute low-back pain is usually counted as an injury, chronic low back pain as a disease. In clinical practice, the great majority of occupational disorders are injuries sustained in industrial or motor vehicle accidents while on the job. (Safety and the prevention of injuries are discussed in Chapter 6.)

Among occupational injuries, the most common site of injury is the back, followed by the hand. Low-back pain is often associated with poor lifting practices and poor workplace design. The prevention of occupational injuries through redesign of the workplace and the control of physical hazards is a major challenge that falls into the domain of two non-medical occupational health professions: ergonomics and safety engineering. Ergonomics is the discipline concerned with designing the workplace and tools to conform to human factors for optimum efficiency, comfort, and safety (see Chapter 9).

#### **Patterns of Occupational Injury on a Large Scale**

Roughly one-fourth of injuries overall are to the back and neck. Back injuries are responsible for over 40 percent of claims, and account for a large number of complaints that go unreported. Second in frequency are hand injuries, accounting for approximately one-fifth of all occupational injuries. The rest are distributed unequally among other parts of the body.

To understand injuries and the pattern of injuries, it is important to understand the calculations used in analysis. It is difficult at first

glance to interpret injury rates. The key indicators are rates (injuries per year or other unit of time) to focus on. The number of injuries that occur in a given industrial sector, is usually based on lost-time claims. These are claims for lost-time injury reported to the workers' compensation carrier. The number of cases indicates the contribution to the overall problem. These figures miss the many minor injuries that do not result in loss of time, however, and are not accurate in reflecting incidence for diseases or for injuries with a chronic or cumulative component. They are reasonably accurate for most injuries, however, and provide a basis for year-to-year comparisons. The absolute number of injuries is of some interest in terms of the contribution of a particular industry to the overall problem, but the rates and the number per employees per year are most useful in setting priorities because these figures identify industries or occupations at highest risk.

Using rates, one can classify various industries into high and low risk compared to the average rate of injuries for all industries in all enterprises in terms of claim rates. High-risk industries would be those whose rate of incidence falls above the average rate; low-risk industries would have a rate below the average. Over time, there has usually been a fair degree of stability in the lost-time claims experience for the various major industrial sectors. Forestry, fishing, and mining usually lead in injury rates; construction, manufacturing, transport, and utilities also rank among the highest-risk industries.

There is much more to the picture than the overall experience of industries as a whole. These rates are aggregate numbers. They are the sum of the experience of a number of smaller trades, industries, and industrial operations within each large industrial sector. Breaking these results down by components of the industry can yield very interesting insights. In the oil and gas industry, exploration has a very high risk of injury, which is diluted by the much lower risk in the production and refining subsectors. For example, another clearly high-risk industry is construction. It is easy to see why construction would be a high-risk industry. It involves many trades and operations that have inherently high risk. The risk profiles of the various trades

within the construction industry are fairly consistently elevated. This represents an industry in which the injury rates are a general problem.

Comparisons are frequently made between large and small industries. The conventional wisdom in occupational health and safety circles is that small enterprises usually have higher injury rates than larger enterprises because they lack the resources to control and monitor hazards. It is generally true that smaller employers are at higher risk of injuries than are larger employers. However, small companies are the norm in many high-risk industries. The company engaged in servicing wells, wildcatting, cutting timber, and performing small-scale construction is likely to be a small employer, usually working under contract to a larger company in the same industry. Other high-risk industries for injuries, such as construction, are dominated by small enterprises. Injury rates therefore reflect employer size in high-risk industries as much as the relative merits of health and safety control in small versus large enterprises.

### **Clinical Management**

This subsection is very general, outlining approaches to occupational injuries that apply regardless of circumstances or body part. The *ACOEM Practice Guidelines (APG)* provides detailed, authoritative guidance on acute injury care and subsequent management, and should be close at hand for every physician who practices OEM or who sees more than the occasional work-related injury. The *APG* establishes best practices, appropriate intensity of treatment, expectations for duration of recovery, standards for return to work, and indicators for change in management when recovery is not proceeding as planned.

The OEM physician must always keep in mind that the capacity to work means more than productivity for the employer and income for the worker. It is also part of the injured worker cum patient's social role, which, if not defining his or her identity, plays a critical role in integration in society, support for the family, self-respect,

fitness and conditioning, and overcoming social isolation. The period following acute treatment for the injury must therefore be managed as carefully as the provision of care. This usually does not happen. The result is suboptimal occupational healthcare, even if the acute-care treatment was superb.

The physician, from the time of first encounter, should establish expectations for recovery. The injured worker is concerned mostly with functional recovery and only secondarily with return to work; however, if return to work is not addressed early on, the omission may convey the wrong impression and shape expectations negatively. In all but the most devastating injuries, it is appropriate to at least mention and, if justified, to discuss in some detail the likely time frame when the injured worker is likely to return to work. This does not have to be a specific date. An indication, however, of whether the injured worker is likely to be off work for hours (returning the same day), days, weeks, or months immediately establishes the temporal framework on the basis of which both the injured worker and the employer can plan.

A subtle change in identity is required. The “patient” needs to become an “injured worker,” temporarily disabled but soon (or not so soon) to return to a social role as breadwinner, co-worker, and full participant in society (a role that is difficult without a job). The patient is dependent, assumes a sick role, and has an uncertain future. The injured worker retains control of his or her future, is only temporarily sidelined, and has a clear pathway back to his or her job and social role.

The history of the injured worker should include the circumstances of the incident. In many encounters, especially in the emergency room, the events leading up to the injury are recorded only briefly or carelessly. This first record is a critical document, however, which if incorrect can throw off all subsequent documentation and compromise the fairness and accuracy of workers’ compensation claims. Cases often hinge on what happened first and what the injured worker said in the first encounter as opposed to later in the investigation. The result of a sloppy or overly abbreviated medical record may be a denial of claims and benefits. It is not necessary to

record the incident in detail, but a short description of what happened, prefaced by “Patient says . . . ,” can be invaluable to the expeditious handling of the case later.

Likewise, filing a “physician’s first report” initiates the workers’ compensation claim and is a critical step in the process of recovery or, in the case of permanent impairment, appropriate compensation. It must be accurate, as complete as possible (notwithstanding the tiny space within which one must usually write on the form), and above all correct. An error or sloppy entry in a physician’s first report can haunt a claim throughout appeals, just as an incorrect entry on the emergency record can. In workers’ compensation, second injuries are handled differently from first injuries. It is important to record whether the patient had sustained a similar injury in the past or had a preexisting impairment, so that the order of injuries will not be confused, as well as the level of impairment that resulted from the first injury, resulting in the baseline for the present event.

Asking the patient to duplicate the action performed at the time of injury can be very helpful. A screening neurological examination should be performed routinely in all cases of major trauma or suspected systemic toxicity, and should include a mental status evaluation. Evidence that the worker was impaired at the time of the injury should be noted, and may trigger alcohol or drug testing; however, this is fraught with legal risk unless the patient’s consent can be obtained or there is a medical reason to obtain these tests.

The injured worker’s authorization to contact the employer is required so that the employer can be notified regarding how long recovery will take and when to expect the worker back. This is critical for planning return to work and can often save the worker’s job. Employers who know that a worker is recovering are more likely to keep the position open and to cooperate with a phased return to work. A phased return to work might involve job accommodation (such as a temporary adjustment in the workplace), temporary part-time, partial-shift work, or limitations on duties (such as lifting). These should be as specific and job-related as possible. A general statement prescribing “light duty” is not sufficient. Light duty

generally means accommodation of some sort (poorly defined) and limitations on tasks, with the expectation that other workers will do the heavy or difficult work. The employer may not know what this means in practice, may think that there is no “light duty” suitable for the worker, and may default to a policy specifying that no employee should return to work before that worker can do everything required by the job. On the other hand, a clear statement of what the worker can and cannot do, kept as short as possible, may allow a phased reentry to the workplace and the opportunity for a better outcome. If an employer does not allow a phased reentry and requires complete functional recovery, it is better for all parties to know this early so that planning and targets for rehabilitation can be adjusted accordingly.

During recovery, OEM physicians often use physical therapy more aggressively and intensively than would be the case in general medical care. There are good reasons for this. One is that physical therapy can be prescribed and formulated to prepare the injured worker for return to work. Tasks and activities that are required at work, such as lifting, dancing, or fine hand movements, can be duplicated in the physical, occupational, or hand therapy clinic. Such activities make the therapy meaningful and relevant, and link progress in rehabilitation to the expectation of return to work. It is also well known that injured workers can become severely deconditioned, to the point of losing strength and agility, when they are off work. Although enthusiasm for “work hardening” has cooled in recent years, physical therapists can design moderate exercise programs in addition to specific rehabilitation to prevent deconditioning and to encourage fitness, weight loss, and retention of the skills or strength that will be needed again on the job. Physical therapy also gives the injured worker a new “job,” one involving scheduled appointments, interaction with others, physical activity, performance evaluation, and goals. When home exercises are involved, physical therapy also puts some responsibility on the injured worker for his or her own care and recovery. While it is not the same as returning to the workplace, adherence to physical therapy appointments and prescriptions is a partial substitute for the discipline and order of a work routine.

Physical therapists also play a very useful role in monitoring the injured worker. In general medicine and surgery, it is not necessary to time the end of rehabilitation precisely, so this is often determined by cost or number of allowed treatments. In OEM, the goal is safe return to work as soon as the injured worker is ready but no sooner. There are risks of re-injury and failure if the worker goes back before the condition is stabilized. Such cases often end in the worker losing employment permanently and experiencing a lifelong setback (even if the worker does not appreciate this at the time). However, extending the recovery time indefinitely exhausts the employer's patience and capacity to readmit the worker to the workplace, and dramatically increases costs for all concerned. Ideally, as the injured worker begins to approach fitness to return to work (discussed in detail in Chapter 18), the physical therapist will contact the OEM physician and recommend an accommodation for the workplace. If the worker can return to most duties at work but cannot do certain tasks, the physical therapist can usually best determine what is safe, what will advance recovery, and what specific accommodation may overcome the remaining temporary disability. Phased reentry into the job, with a period of accommodation or part-time (such as half-day) work, helps with resocialization and reintegration into the workplace community, as well as with restoring the job demands from the real world that condition the body during the next and final stage of recovery.

The injured worker often expects to be off work longer than is necessary, if not informed otherwise. It is therefore wise for the treating physician, as well as the physical therapist, to refer at every visit to the expected time frame and to count down from the time the injured worker is expected to return to work. The OEM physician may then propose a date for return to work, and may suggest whether it should be a full day or temporarily part-time. Because of the natural human tendency to want additional time, the patient should not be asked if he or she feels ready to go back to work. The terms for returning to work should be proposed by the physician, and if the injured worker raises a valid reason to delay, this should be discussed.

There is a natural tendency, even among highly motivated workers, to mope when they are injured and to hold back from their usual work and personal routine. There is an old stereotype in workers' compensation stories of injured workers spending their days in front of the television watching soap operas. It is important for full recovery and to prevent depression that the injured worker do as much as possible and not withdraw and vegetate. The worker's family, support network, and community will usually sustain the worker in his or her accustomed social role but will take their cues from the injured worker, enabling the sick role if the worker acts dependent.

Injured workers should be encouraged, within limits of safety, to push the envelope and to do things around the house and in the community that keep them active, engaged, and striving to expand the limits of their capacity. It is prudent that the OEM physician or physical therapist document what the injured worker is capable of doing, however, because sometimes workers' compensation carriers (insurance companies) use surveillance methods to capture injured workers in the act of doing something strenuous or inconsistent with their impairment. It is then important to determine what was medically justified and what was not.

One intervention that does not seem to work well is injured worker support groups. This strategy has been tried in various forms and has proven disappointing in practice. Participation is by self-selection and attracts those for whom such groups merely serve to reinforce a sense of victimization.

### **Injury on the Scale of a Workplace**

Table 16.1 presents the body parts affected by injuries, in the same data sets, demonstrating the distribution of injuries by type. Clearly, the majority of injuries fall in the category of strains and sprains. This is consistent with other data for occupational injuries that show soft-tissue injuries to be much more common in occupational practice than serious lacerations or fractures. After the back, it should be noted the hand is the body part most often affected.

## Occupational Injuries and Musculoskeletal Disorders

**Table 16.1.** Distribution of Disabling Nonfatal Work Injuries and Illness, California, 1988 ( $n = 371,738$ )

	Percentage of Total	Body Part Most Often Affected
Strains, sprains, hernias	42.9	Back (48 percent), lower extremities (18%)
Cuts, lacerations, punctures	15.6	Upper extremities (69%)
Fractures	10.9	Upper extremities (42%), lower extremities (38%)
Contusions, crushes, bruises	10.8	Lower extremities (37%), upper extremities (28%)
Abrasions	4.3	Eye (81%)
Burns and scalds	2.6	Upper extremities (45%)
Concussions	0.8	Head (100% by definition)
Multiple injuries	0.6	—
Amputations	0.2	Upper extremities (97%)
Electrocution	0.1	—
Other	11.0	—

Statistics kept by the Occupational Safety and Health Administration show a relentless decline in reported injuries over the last two decades, with the implied message that control efforts are working. Many occupational health professionals think that the true trend is not so rosy, pointing to evidence (anecdotal or statistical) indicating that relatively minor injuries are underreported. It is also claimed that as much as 40 percent of injuries in California, a state to be commended for acknowledging and quantifying the problem, go unreported. Penalties and workers' compensation insurance premiums do create

incentives to minimize recognition of workplace injuries, to return workers to work immediately (so as not to incur a reportable injury), and to discourage workers from reporting minor injuries. The result, often, is the rather unbelievable claim of “zero days lost to work injury” over a period of several years. The reduction in the injury rate that has been reported may be due as much to the changing profile of industry as to any workplace intervention, since there are fewer occupations in high-risk manufacturing jobs as the economy in North America shifts to services and higher value-added technology.

Table 16.2 illustrates the use of statistical data for injury tracking, using data from California from the late 1980s, when reporting was probably more accurate. Crude data are not very helpful, but rates per person-year of employment are quite revealing. It can be seen that injury rates vary greatly by industry sector, but that rates within sectors are more or less stable year to year. It is also obvious that construction, a sector with a consistently high rate of injuries worldwide, led all sectors for injury risk in the state in that period, given the sectors active in its economy and their prevailing technology.

Several occupations in North America continue to be unacceptably dangerous, even taking reports at face value. These include the traditionally high-risk occupations of construction, underground mining, fishing, and tree cutting. Construction is perennially high in risk, both for injuries as a whole and for fatalities, in part because much of the workforce is transient, there are many hazards, and shortcuts are often taken on safety measures. Additionally, in the United States, recent immigrants are disproportionately represented, and for those who do not speak fluent English, training is often not available in the first language of the worker (especially if it is not Spanish).

There is evidence that high-risk occupations are concentrated in lower-paying industry subsectors that are less closely monitored. These tend to be aggregated into larger sectors where their statistical excess is less obvious, and in geographic locations where they are not as visible. Even before the current recession began, the newly restructuring economy failed to deliver its promise of clean work and huge leaps in productivity. Instead, there has been an intensification of economic

**Table 16.2.** Lost-Time Claim Rates by Major Industry Sector, California, 1987 to 1988

	1987	1988	Change
<i>Agriculture, forestry, and fishing</i>			
Person-years <sup>1</sup>	407	428	+5.1%
Lost-time claims <sup>2</sup>	19,728	21,752	+10.3%
Lost-time claim rate <sup>3</sup>	4.8	5.1	+1.1%
<i>Construction</i>			
Person-years <sup>1</sup>	575	607	+5.6%
Lost-time claims <sup>2</sup>	48,636	51,906	+6.7%
Lost-time claim rate <sup>3</sup>	8.5	8.6	+1.0%
<i>Manufacturing</i>			
Person-years <sup>1</sup>	2,108	2,149	+1.9%
Lost-time claims <sup>2</sup>	74,366	81,492	+9.6%
Lost-time claim rate <sup>3</sup>	3.5	3.8	+1.1%
<i>Transportation and Utilities</i>			
Person-years <sup>1</sup>	541	548	+1.3%
Lost-time claims <sup>2</sup>	24,296	27,920	+14.9%
Lost-time claim rate <sup>3</sup>	4.5	5.1	+1.1%
<i>Trade</i>			
Person-years <sup>1</sup>	2,072	2,150	+3.7%
Lost-time claims <sup>2</sup>	68,112	73,830	+8.4%
Lost-time claim rate <sup>3</sup>	3.3	3.4	+1.0%
<i>Public administration</i>			
Person-years <sup>1</sup>	1,534	1,580	+3.0%
Lost-time claims <sup>2</sup>	62,198	67,752	+8.9%
Lost-time claim rate <sup>3</sup>	4.1	4.3	+1.0%
<i>Mining</i>			
Person-years <sup>1</sup>	41	42	+2.9%
Lost-time claims <sup>2</sup>	1,728	1,690	-2.0%
Lost-time claim rate <sup>3</sup>	4.2	4.0	-1.0%
<i>Services</i>			
Person-years <sup>1</sup>	2,994	3,143	+5.0%
Lost-time claims <sup>2</sup>	69,296	75,372	+8.8%
Lost-time claim rate <sup>3</sup>	2.3	2.4	+1.0%

1. Employment in thousands.

2. Actual count of injuries and illnesses.

3. (2) divided by (1).

Source: 1988 California Work Injuries and Illnesses. San Francisco, California Dept. of Industrial Relations, Division of Labor Statistics and Research, July 1989, p.14.

pressures on marginal workers. The effects on employment include increasing income disparities, fewer jobs that lead to viable career tracks, less secure employment, longer working hours, and higher risk of uncompensated unemployment. Workers who are either newly entering the workforce or who do not have the training or skills to trade up in employment are under pressure from falling middle- and lower-income wage scales and increasing income disparities. Their opportunities are increasingly limited, whether they are seeking better-paying jobs with other employers or moving up the now-shortened ladder of the same employer. They face a declining choice of employment opportunities and well-paying jobs; those jobs that do exist tend to be concentrated in residual dangerous work in manufacturing or service subsectors that are often small or economically marginal.

These economic and social trends appear to be pushing workers into choosing among the diminishing but still substantial number of dirty and dangerous jobs that are available, as illustrated by the extreme example of McWane Pipe, a steel pipe manufacturer that was featured in a 2003 series in the *New York Times* as America's most dangerous employer. McWane manufactured a metal pipe used locally in oilfield work and other, less valuable products. The economics of the product kept the company from moving offshore, since the transportation costs for the pipe would raise the cost above that of local production; however, to maintain margins, the company kept wages down. The workers at the company were generally poorly skilled and unable to relocate. They accepted the work despite the risk because although it paid poorly on a national scale, it was better paying than alternative jobs for which they could compete. The high number of injuries, including fatalities, at this one company was easy to lose track of in the context of the steel manufacturing sector as a whole, where the risk is much lower.

Those workers who are forced into these dangerous jobs are, predictably, the poorly prepared, the new entrant, the recent immigrant, and the illiterate worker. National occupational injury statistics may conceal the experience of these marginal industries by aggregating them into much larger economic sectors.

### ***Fatalities***

Fatalities due to injuries on the job are tragedies affecting all concerned and society as a whole. Approximately 7,000 workers die each year on the job in the United States, almost all of them (94 percent) men, mostly ages 20 to 35. A death on the job has profound repercussions for the family, the community of which the worker was a part, and co-workers. A fatality usually strains and sometimes breaks the bonds of trust between management and the workers, and often leads to ruminative brooding, blaming, denial, and scapegoating.

Because of their extreme nature and implications, fatalities are handled differently from other injuries by occupational health and safety authorities. Fatalities are counted separately, and trigger mandatory investigations. Compensation to the family is usually a fixed settlement. A fatality often results in an extended process of appeal and litigation and the psychological consequences of compensation-dependency behavior.

Fatalities that occur suddenly from an incident on the job are the most wrenching and dramatic types of occupational injury but are not the only devastating events in the workplace. Deaths from occupational illnesses, such as cancer or lung diseases, are more difficult to document and usually occur late in life. They are often unrecognized and may be more common than believed. Severe injuries resulting in permanent disability may be as emotionally devastating and costly to the system and almost as disruptive to the community as traumatic fatalities. Accompanying sudden fatalities and seriously disfiguring or disabling trauma is a natural response to the emotional reaction associated with a death, and is a way of objectifying and dealing with the shock, grief, and sense of responsibility felt by the survivors.

Fatality rates (usually expressed as deaths per 100,000 workers per year) are highest in: mining (31.9); transportation, communication, and public utilities (25.4); construction (24.0); and agriculture, forestry, and fishing (20.7). The transportation industry contributes the majority of fatalities in the transportation, communications, and public utilities sector; the individual components of the agriculture,

forestry, and fishing sector are all of high risk. After these four highest-risk sectors, there is a big gap to the other, lower-risk industry sectors, all of which have fatality rates below 10. The lowest risk is in wholesale trade. Individual occupations at greatest risk include transportation workers, farmers, laborers, and workers in trades and crafts (other than machinists). The lowest rates are found among clerical workers.

The distribution of fatality risk largely parallels patterns of employment in the occupations at highest risk. African Americans have a slightly higher risk than whites. Alaska and Wyoming have had the highest rates. Although the greatest proportion of deaths occur in younger age groups, the highest risk is sustained by workers above age 65 and particularly age 70 and older.

Certain patterns of fatality are recognized and can be addressed relatively easily. Homicides are responsible for 11 percent of fatalities among men and almost 40 percent among women; homicide is a particular problem in retail services, and is usually associated with robbery and sometimes with attempted rape. Security measures to ensure visibility, ease of communication, and backup for convenience store personnel at night have been instituted by major chains to address this problem. Multiple fatalities commonly occur in the context of a chemical release, fire, or confined space incident. Once a worker (usually a male) is down, his co-workers often forget or ignore their training and attempt a rescue without obtaining proper personal protective equipment. All too often, the result is two fatalities (or more) and a fatal time delay in what otherwise might have been a successful rescue of the first victim.

Construction sites are particularly hazardous, and many fatalities are associated with improper trench-digging procedures and failure to shore up the sides of the excavation. Cave-ins are common in such situations and can be lethal when they trap workers. Even among survivors, the asphyxiation associated with burial in sand or wet soil obstructing the airway and compressing the thorax may lead to oxygen deprivation and permanent brain damage before rescue. Prompt rescue is obviously critical, but even if rescue is delayed there may be hope of an air pocket or a breathing hole sufficient to sustain life.

### ***Incident Investigation***

Once a major incident resulting in injury or major property loss occurs, especially if a fatality is involved, an investigation is usually conducted. A team is assembled of persons with technical knowledge of the situation, occupational health and safety expertise, and awareness of management responsibilities. An OEM physician is typically included on such teams.

The usual methods of conducting incident investigations include site visits, reconstructions and simulations, and interviews. The area should be cordoned off and locked up or guarded if possible. Work is typically halted in the workplace until the investigation is completed and the workplace is judged safe again.

The site of an incident should be visited as soon as possible, before the scene is disturbed and problems are corrected. There are often motives to tamper with the evidence, especially if negligence is a possible factor or if there is an attitude of distrust. Investigators should be personally equipped with appropriate protection to ensure that they are not at risk. Photographs and a tape measure are essential aids to reconstructing the physical environment, and drawings and diagrams should be drawn freely as needed.

Interviews are often difficult to conduct because the atmosphere may be fearful, hostile, or passive. Witnesses often influence one another's perceptions and may even come to a "consensus version" through discussion of the case among themselves. They may try to protect or blame a party they feel is culpable, or they may rationalize their own actions. Often, their motives to distort the narrative are unconscious and confused, and may reflect distrust of the employer or a perception of responsibility or morality that is not necessarily the same as legal liability or objective certainty. As in courtrooms, eyewitnesses are not always reliable.

Establishing a rapport with the witness is critical to the success of such an interview. The attitude of the investigator should always be polite, objective, and professional but concerned. The interview should be conducted, whenever possible, in private with only one

or a few interviewers. The witness should be put at ease and reassured that the intention is not to assign blame. The narrative the witness gives should be listened to carefully on first telling, without interruption or distraction; it can be recorded (for transcription later), but the interviewer should make eye contact and pay close attention rather than writing notes. The witness should then be asked to repeat the sequence of events, answering questions designed to fill in missing information. The questions should be simple and direct, but open-ended and not leading. As the interview proceeds, the questions may become more specific and detailed, but should never appear to entrap the witness or lay blame. It is usually wise to interview key witnesses twice, after they have had a chance to jog their recollection. Witnesses should be asked to call back if they remember anything else of significance.

Note taking should be unobtrusive but not hidden. The witness should be made to feel that the notes are intended to highlight pertinent details of the incident and not to record their reactions or comments. It is useful to review one's notes with the witness at the end of the interview to ensure that the notes are correct and complete.

The framework for incident reconstruction should be Haddon's matrix (discussed in Chapter 6). This framework has the great advantage of placing emphasis on the elements of the incident and not on the people, and so is less threatening than an interview that focuses only on a chronological stepwise reconstruction of events. Particular emphasis should be placed on the pre-incident factors: unsafe acts by the injured party and others, personal factors, environmental conditions, ergonomic factors, training deficiencies, management responsibility, timing and shift, and the experience of the injured party and others on the team. Consultants should be used as needed to pinpoint the significance of forensic evidence or unfamiliar hazards.

Psychological aspects of the incident are usually the most difficult to reconstruct. The perception of a worker's attitude at the time is often distorted by past experience and reputation. Co-

workers may project their own feelings and motives onto the injured party. The conduct of an unsafe act, failure to wear protective equipment, and deviation from established procedures should be treated as factors to be satisfactorily explained, not as basic causes in themselves.

Reconstruction of the incident usually begins by establishing the chronological sequence of events, proceeds with the identification of any unusual or particularly hazardous preexisting conditions, seeks to explain individual behaviors, and ends with an attempt to identify solutions to the problems identified.

Detailed reconstruction of the incident often yields valuable insights into how well or how poorly the work is conducted and safe procedures are followed. Together with information on past trends of injury and property loss, such an effort may suggest needed improvements in procedures, training, and hazard control.

Corrective action suggested by the investigation should be prompt and visible. It should be monitored for effectiveness in solving the perceived problem. Any deviations from recommendations of the investigating team should be explained satisfactorily.

To be most useful, the conclusions of the investigation should be widely available to workers, unions, management, regulatory agencies, and insurance carriers. An open investigation results in confidence in the conclusions and is more likely to result in cooperation to solve the problem. Concealment or suppression of the findings is an invitation to allegations of cupidity and negligence. Openness builds trust.

Throughout the investigation, contact should be maintained with the injured or affected worker and with the surviving family members and co-workers of a fatality. To further their own healing, they should not be forgotten and left to wonder what is being done. Knowledge of the circumstances surrounding the incident contributes to the grieving process and is an important part of rehabilitation. Some workers and families may require psychological counseling for posttraumatic stress syndrome. All require acknowledgment of the legitimacy of their emotional reactions.

### ***Burns***

The medical literature on occupation-related burns emphasizes treatment rather than prevention. Several surveillance systems and studies of occupational injuries have reported on the frequency of burns, but few have characterized the victim, the workplace, and its risk factors or the circumstances of the injury. Burns are associated with a higher risk of permanent disability than most other types of occupational injury.

At one regional burn treatment center (in San Diego, California), six different types of burns were identified, representing all major recognized types except those caused by molten metal. The three deaths were in two burn categories. Two deaths resulted from flame burns and one resulted from an electrical burn.

Scalds were the most common burn injury overall (38 percent). Scald burns had the lowest disability rate at 6.5 percent.

Flame-related burns resulted in the most hospital admissions, were more common for men than for women, and resulted in the highest average percentage of body surface area burned. Forty-seven percent of the hospitalized cases were for flame burns. Flame burns had the highest mean and median number of days off work (58.7 and 39 days). Individual cases of flame burns also had the longest hospital stays.

Electrical burns were the most likely to be associated with permanent disability. Electrical burns made up only 11 percent of the hospitalized cases, but 85.7 percent of the victims were disabled from their burn injuries. Electrical burns had the highest mean and median number of days hospitalized (25.8 and 25 days, respectively). As a group, the electrical burn patients required the most surgery, with fully 77 percent of them requiring general anesthesia for skin grafting or amputations.

Tar burns and scalds due to hot grease emerged as surprisingly common injuries. Chemical burns were uncommon but often severe. (See also Chapter 17 for acute skin injury.) Contact burns were more often observed among younger persons, and electrical burns among older patients. The other rates of disability were 14.3 percent of the tar burns, 50 percent of the chemical burns, and 25 percent of the

contact burns. Tar-related, electrical, flame-related, and chemical burns affected men almost exclusively.

Occupational associations showed few surprises. Among semi-skilled and unskilled workers, roofing workers sustained the most tar burns. Service workers, primarily in the restaurant industry, almost exclusively experienced scald burns; 55 percent of these were caused by hot grease. Among skilled workers, firefighters sustained flame-induced burns, and electricians sustained electrical burns. Compared with their representation in the general population, semi-skilled and unskilled workers were greatly overrepresented, and professional and technical workers were greatly underrepresented.

Representative incidents included electrical burns in farm workers, grease melting through plastic containers in fast-food restaurants, a bucket of solvent ignited by sparks from arc welding, facial burns from a flashback in a tar kettle, and fingers caught in a laundry press.

The typical burn patient was a white male aged 21 to 30 who suffered a scald or flame-related burn while working as a semi-skilled or unskilled laborer or restaurant worker, and who was admitted to the hospital for about two weeks but was not permanently disabled.

Occupation-related burns often result in disability, and in some cases death. Virtually all can be prevented. Despite decades of apparent improvement in reducing the incidence of burn injuries, progress appears to be slowing. Further improvement is likely to depend on a more sophisticated approach to burn prevention, based on known characteristics of the worker at risk, the hazard, and the workplace environment. This is no easy task. It is a considerable professional challenge for safety personnel even in the relatively controlled environment of a large employer, but it becomes daunting indeed in the context of small, scattered workplaces such as restaurants, roofing sites, laundries, and farms.

Worker education is desirable, but efforts are easily rendered ineffective by language problems, turnover of personnel, and negative attitudes. Changes in personal protection and equipment design are badly needed, especially in high-risk occupations, but their adoption is complicated by the widespread nature of the problem and the pres-

ence of thermal, electrical, and chemical hazards in virtually every building, workplace, and home.

## MUSCULOSKELETAL DISORDERS

Musculoskeletal disorders, as noted, are conditions that, with the exception of acute low-back pain, are slow to develop and chronic in nature, rather than injuries. They are the product of an imbalance of forces rather than a single traumatic force, which may initiate acute back pain. Because of this, it is reasonable to think of musculoskeletal disorders as a class of disorder that represents the consequences of ergonomic failure (see Chapter 9). The *APG* is an authoritative guide to evaluating and treating these conditions.

In 2001, an ergonomic standard developed by OSHA after ten years of research and preparation was stopped by an act of Congress, which usually does not concern itself with individual regulatory standards. The standard would have required employers to identify and control hazards that could result in musculoskeletal disorders, to develop ergonomic programs for worker protection, to collect data on such disorders claimed in their workplaces, and to provide medical care for injured workers. The stated reason for this defeat was that the scientific basis for an ergonomic standard was weak. However, authoritative bodies such as the National Academy of Sciences had concluded that the science behind ergonomic assessment and the etiology of musculoskeletal disease was sound. That does not mean that the literature is perfect, but it is sufficient. More research is certainly required, especially with respect to the efficacy of interventions, but there is enough to act on.

Human beings are more complicated than biomechanical models. They have inapparent mechanical weaknesses, histories of prior conditioning or de-conditioning, life situations, psychological issues that affect perception of pain, and different responses to stress. This is particularly apparent for chronic low-back pain. It would be extraordinary if other factors did not aggravate, amplify, or modify the effect of ergonomic factors. However, the presence of other factors that

may aggravate, amplify, or modify the effect of ergonomic factors does not obviate the conclusion that mechanical or ergonomic issues are primary among risk factors.

Non-ergonomic risk factors and influences that modify symptoms and disability should be taken into account in prevention and intervention programs. However, non-ergonomic risk factors should not be confused with the underlying cause of musculoskeletal disorders. The existence of non-ergonomic risk factors must not be used to deny the cause-and-effect relationship between ergonomic risk factors and musculoskeletal disorders. If ergonomics is not primary, the result will be ineffective prevention and cynicism on the part of workers.

The debate has been cast as an issue of working versus not working, or pain arising out of work versus discomfort reflecting the activities of daily life. This is a distortion of the issue. The real issue is working in ways that are safe, adaptable, and efficient. Modifying work practices to protect the worker from the risk of musculoskeletal disorders has the benefit of making such work more efficient and sustainable. Therefore there will be productivity gains from better ergonomic design. (See Chapter 9.)

Low-back pain and upper extremity disorders should be considered separately. Upper extremity disorders are clearly associated with ergonomic factors, but the response of the individual may vary. Low-back pain is more complicated, but ergonomic factors clearly play a substantial role.

The risk of musculoskeletal disorders of the upper extremity is concentrated in certain occupations characterized by identifiable risk factors related to the repeated movement against resistance of an intrinsically weak joint or previously injured body part; this risk is directly related to the frequency and force of such movement. This is not to say that other factors are not involved. However, the primacy of the ergonomic risk factors is evident in upper extremity musculoskeletal disorders.

Low-back pain is more complicated. While it is absolutely clear that a variety of situational factors affect the risk, perception, and

disability associated with musculoskeletal pain, there is also sufficient evidence to conclude that ergonomic factors play a role.

Not everyone becomes disabled from chronic low-back pain. Some people do learn to adapt or persevere, especially if they must continue working to support their families. That does not mean that their pain is less, that their lives are unchanged, or that they are less worthy of protection. Unemployment and time off work to recover from a disability can be a profoundly corrosive and demoralizing experience. It can cause its own psychological injury and disability. However, the corrective is not to force people with pain to work despite their discomfort and suffering and to deny their right to a safe workplace.

Reasonable accommodation, and the technology it brings, make the workplace more adaptable, productive, and flexible, and provide opportunity to expand the workforce. An ergonomically sound workplace is usually safer from the risk of traumatic injury as well, and is always a more efficient workplace. (See Chapter 9.) Ergonomic interventions have the objective of making work more biomechanically efficient, so that workers are able to continue their duties with less strain, less energy expenditure, greater mechanical efficiency, and therefore greater stamina and duration.

Modifying work practices to protect the worker from the risk of musculoskeletal disorders has the benefit of making such work more efficient and sustainable. Therefore there will likely be productivity gains. As in other areas of occupational health, the changes that are effective in preventing musculoskeletal disorders are likely to make work safer, more efficient, and more productive.

### **Low-Back Pain**

Low-back pain (sometimes called “lumbago”) is the most common occupational disorder characterized as a “disease.” It is a source of continuing concern because of the prolonged recovery and time off work that it imposes on a substantial percentage of patients. Low-back pain is a nonspecific description of a pain syndrome associated with

soft-tissue injury rather than a particular diagnosis; few cases are demonstrably associated with disease, and very few with spinal fractures or bony lesions. It is often associated with lifting objects but is not necessarily determined by the weight of the object, and it can be precipitated by awkward posture. It is most often a problem of men between the ages of 40 and 70. Identification rests on subjective symptom reporting, which is easily influenced by mood and anxiety. No diagnostic maneuver is consistently informative. Management is paradoxical because almost every intervention seems to help to some degree, but almost nothing is definitively curative in most cases. The association with work in a given case may be exceedingly difficult to determine because the disorder is common in daily life and subject to aggravation on the job.

Low-back pain affects approximately 85 percent of North Americans at some time during their lives, and 2 percent of workers in industry per year in the United States. The costs to industry associated with these cases has been estimated at \$25 billion per year, almost half of which is distributed more or less equally (around 15 percent) among physical therapy, inpatient care, and medication costs. Most of this enormous cost, about 80 percent, results from a small percentage of cases (approximately 10 percent) in which low-back pain becomes chronic, and that are characterized by a combination of delayed return or failure to return to work, chronic pain, psychological adjustment problems, functional impairment, surgery with poor results, and litigation.

The essential rule in managing low-back pain, emphasized by the *APG* and numerous other guidelines and reviews, can be summarized as "First do no harm." If no major pathology is demonstrable, the patient should always be treated conservatively, with mild analgesics (acetaminophen is usually recommended), only short periods of bed rest (two days maximum is usually recommended), and a return to activity as soon as possible. Great harm can be done by giving strong, dependency-inducing analgesics for a condition that does not require such treatment and usually resolves quickly. Prolonged bed rest (as long as six weeks was recommended in years

past) and inactivity severely decondition the injured worker and perpetuate the back pain.

The great majority of cases of low-back pain recover quickly, within days. Patients who do not may progress within weeks to chronic back pain, if they have not had it before. Most will learn to tolerate occasional twinges and backaches. Others enter a new phase in which dependency, fear of pain (and the desire to be completely free of any pain), the chronic pain syndrome, deconditioning, and separation from work routine combine to make recovery and return to work increasingly problematic. Workers' compensation data suggest that this transition point is at about three months, after which injured workers who have filed claims (many do not, especially for mild symptoms) are unlikely to return to work at all. This group, not surprisingly, incurs the major cost of benefits for low-back pain (on the order of 80 percent) although they make up a relatively small proportion of total claims (about 10 percent).

Before assuming that low-back pain is of the usual benign variety, the physician should rule out major pathology by looking for what are called "red flags" in the history and physical examination. These are uncommon but critical. Table 16.3 presents a list of clinical "red flags" which, if present, indicate potentially serious disease that must be ruled out. Signs of compression of sacral roots, such as diminished cremasteric or anal reflexes, anesthesia in a "saddle" pattern, and a history of autonomic dysfunction in the pelvis, constitute the "cauda equina syndrome," a medical emergency requiring immediate neurosurgical consultation. A history suggesting vertebral fracture (recent trauma, corticosteroid use, age over 70 years), signs suggestive of cancer (history, unexplained weight loss, age over 50), signs suggestive of infection (history, fever, urinary tract symptoms, immunosuppression, intravenous drug use), and signs of abdominal aortic aneurysm (pulsatile abdominal mass, pain at night and at rest) constitute evidence of major pathology that must be investigated, and certain laboratory tests may be in order: complete blood count, sedimentation rate, and urinalysis, for example. Pelvic disease can also present as low-back pain. Only

**Table 16.3.** “Red Flags” Indicating Possible Major Pathology in Low-Back Pain

“Red Flag” Finding	Rule Out:
<i>History</i>	
History of axial compression, trauma	Spine fracture (compression fracture) Rule out osteoporosis
Pain when lying supine and at night	Malignancy, infection
Fever, history of autoimmune disorder	Connective tissue disease (lupus, rheumatoid arthritis, ankylosing spondylitis)
History of IV drug abuse	Osteomyelitis of spine, bacterial endocarditis
Immunosuppression or compromised immunity (including HIV/AIDS)	Infection, malignancy
Positive TB screening or history of TB	Spinal tuberculosis (Pott’s disease)
Bowel or bladder incontinence lasting more than a month	Spinal stenosis (cauda equina syndrome), tumor, compression of cord
STD, urethritis, uveitis	Reiter’s syndrome
Arthritis	Reiter’s syndrome, psoriasis, rheumatoid arthritis
<i>Physical examination</i>	
Colicky flank pain, hematuria	Kidney or ureter stone
Abdominal mass (pulsatile) and persistent severe back pain	Abdominal aortic aneurysm, dissecting
Fever	Infection (disc space, spinal tuberculosis, paraspinal abscess), bacterial endocarditis
Neurogenic bowel or bladder incontinence, impaired sacral nerve root reflexes, “saddle” anesthesia (perianal and perineal numbness), lower extremity weakness	Spinal stenosis (cauda equina syndrome)

(Continued)

**Table 16.3.** (*Continued*)

Positive straight leg raising (nonspecific)	Sciatica, nerve root irritation (possible disc herniation); may not change management plan
Localized pain over spine	Fracture, lytic lesion
Pain with spine extension rather than flexion	Spondylolysis (anatomical defect of pars interarticularis)
Inability to squat or walk heel-to-toe	Spinal stenosis (cauda equina syndrome)
<i>Laboratory</i>	
Hypercalcemia	Hyperparathyroidism
Elevated sedimentation rate	Autoimmune disorder, vasculitis, osteomyelitis,
Lytic lesion in bony structure on x-ray (may be elsewhere)	Multiple myeloma, metastatic cancer

about 10 percent of occupational low-back pain cases have evidence of an abnormality on x-ray, and most structural abnormalities on x-ray do not correlate with symptoms, so x-rays are not routinely needed unless there is a specific clinical indication. (Spine films also deliver a relatively heavy dose of radiation and so are to be discouraged unless needed.)

Other causes of low-back pain are uncommon but must be considered if any aspect of the history is unusual or suggestive. Spinal stenosis is an uncommon disorder that presents clinically as signs of lumbar nerve root irritation bilaterally, with pain relieved on flexion of the spine as if bending to touch the toes; it is exceptional among causes of low-back pain in that it is surgically treatable and responds poorly to conservative treatment. Other, less common causes of low-back pain include sacroiliitis (sometimes associated with ankylosing spondylitis and often presenting as buttock or leg pain), pathologic fractures of the vertebrae (associated with malignancy,

bone disorders, or hyperparathyroidism), disc or vertebral infections. Exaggerated low-back pain is often associated with the somatization of depression.

The clinical evaluation of low-back pain in the absence of red flags is usually straightforward. When an acute low-back injury occurs, or when the low-back pain begins to express itself chronically after a time of intermittent discomfort or repetitive strain, the evaluation seldom uncovers major pathology. Most patients present with generalized pain, and a minority (one in four) with referred pain to the buttock or sciatica. A routine clinical examination of the spine would include examination for gait, posture, rotation, scoliosis, and excessive kyphosis or lordosis; diminished mobility; focal tenderness around the spine; palpation for muscle spasm; and reflexes for L<sub>4</sub> (knee jerk) and S<sub>1</sub> (ankle jerk). It would also include standard straight leg raising for evidence of sciatic nerve irritation, testing strength and motor function in the lower extremities, and mapping of sensory abnormalities of a dermatome distribution: L<sub>4</sub> (medial foot), L<sub>5</sub> (top of foot), and S<sub>1</sub> (lateral foot).

The clinical management of acute low-back pain rarely rests on a specific anatomic diagnosis. The medical approach that has emerged over the last three decades is grounded on very conservative and robust principles, mostly codified and validated by the federal agency currently known as the Agency for Healthcare Quality and Research. Some of its guidelines are the following:

- Evaluate the injured worker for the presence of one of several “red flags,” which are symptoms, signs, and findings that suggest major pathology.
- Treat patients with sciatica like other patients with nonspecific back pain; the great majority will recover without special intervention.
- In the absence of major pathology, treat low-back pain as conservatively as possible, with over-the-counter medication (including acetaminophen or NSAIDs).

- Do not use strong analgesics, especially opiates. The weight of evidence suggests that they have no benefit beyond mild analgesics and there is a strong risk of dependency and side effects.
- Do not use muscle relaxants, except possibly to interrupt a muscle spasm. The weight of evidence suggests that they have no benefit beyond mild analgesics and there is a strong risk of dependency and side effects.
- Mobilize the patient as quickly as possible, preferably within two or three days. Prohibit prolonged bed rest.
- Try simple physical modalities if they work in a particular patient: heat and cold application. However, diathermy, TENS, biofeedback, acupuncture, ultrasound, cutaneous laser treatment, stretching exercises, and massage have no demonstrated benefit in acute low-back pain.
- Injection is not appropriate in acute low-back pain.
- Identify contributing causes (obesity, lack of exercise, and deconditioning) and work with the patient to remedy them.
- Avoid x-rayng the spine unless the patient is over 50 and there is a clinical indication.

Prolonged rest is extremely destructive and should be discouraged for acute low-back pain, even at the expense of transient discomfort to the patient. The patient should be encouraged to return to normal activity within two days and to work as soon as possible, even if some symptoms persist. The longer the patient is out of his or her normal routine, the more likely it is that he or she will experience deconditioning, with predictable loss of strength.

When short-term treatment is ineffective and chronic pain persists, the problem becomes much more complicated and psychological factors play a critical role in management. Chronic pain is fundamentally different in mechanism and affective response from acute pain and is discussed in some detail later in this chapter. After a period, variously estimated at around three months, the pain itself

becomes the focus of the disorder and the principal management problem. Thus, the management of chronic back pain is largely the management of chronic pain that happens to be localized in the back. A particular priority in such cases is the earliest possible return to work and meaningful life activity. This may be resisted out of fear, but nonetheless it represents perhaps the single most important step to be taken. Return to work in this context is a primary mode of treatment, not solely an outcome of appropriate treatment.

Chronic low-back pain, on the other hand, is a diagnostic and management conundrum. The relative success of non-specific measures for low-back pain, compared to the inadequacy of more aggressive measures and the risk of complication, has led to a nihilistic view on the part of many clinicians and especially insurance carriers. There are currently two major schools of thought. The older school of thought emphasizes biomechanical factors, and within this school of thought are many ideas about how the spine works and what specific lesions may occur. The other school of thought emphasizes non-physical, particularly psychosocial factors and views low-back pain as primarily an adjustment problem. There are few examples in medicine of such polar differences among mainstream practitioners.

In the biomechanical model, low-back pain is conceptually simple to imagine but difficult to model or understand biomechanically. The human back is doubly curved to accommodate bipedal erect posture. The lumbar lordosis, or inward curve, does not appear until the child learns to stand and walk. The architecture of the spine is well suited mechanically and hydraulically (as a result of the semi-fluid disc nuclei, which serve as shock absorbers) to absorb compression stresses. It is less well suited to accommodate shear stresses and relies on relatively weak interfacetal joints and soft-tissue structures, specifically ligaments, intervertebral muscles, paraspinal muscles, and the bracing effect of abdominal guarding. Forward bending places the back at a mechanical disadvantage because even without any load a shear force vector is applied to the lower back; the weight of the load is less important than the posture and nature of the movement. When loads are lifted in such a way as to place shear stress on the spine, as

with forward bending, considerable force is put on these relatively weak structures. Sudden or abrupt movements are particularly difficult because they are associated with muscle spasms and guarding. Repetitive movement may stress the back structures repeatedly and cumulatively, before the soft tissues have an opportunity to recover fully. Despite this relatively simple impression, the actual biomechanical events that take place are difficult to identify, and the spine (particularly in its soft-tissue structures) is notoriously difficult to model when in motion. As a result, there are many controversies and different schools of thought on the precise mechanics of low-back injury and the optimal approach to back care and injury prevention. The discrepancies relate particularly to the importance of the “pelvic tilt” and to the specific soft tissue structures thought to be involved, particularly the intervertebral disc.

In recent years, opposing viewpoints have placed heavy emphasis on psychosocial factors and treat low-back pain as a symptom of problems related to adjustment, workplace issues, and stress-related strain. The prevalence of low-back pain in the workplace does reflect psychosocial issues, and individual risk for low-back pain is predicted by employee job satisfaction and other stress factors, as well as smoking habits, since smokers are more likely to develop chronic low-back pain than nonsmokers. It is clear that the major risk factors for poor outcome in back pain are psychosocial, especially after 90 days and for men older than 50. Such factors include inability to return to work with the pre-injury employer, complaints that treatment made the condition worse, low job satisfaction, substance abuse, lack of education, frequency of missed appointments, and litigation.

Because chronic low-back pain without major pathology or sciatica is entirely subjective, it is subject to the many distorted perceptions described in Chapter 15. In individual cases there may be an element of symptom magnification, symptom exaggeration, somatization, or sometimes malingering. Occasionally workers close to retirement, for example, may begin to complain of low-back pain, which is very common anyway, under circumstances that suggest a

desire for additional pension benefits for disability. An injured worker may not feel that his or her pain is adequately appreciated, and therefore will express his or her discomfort in an exaggerated manner. There may be outright malingering. A characteristic of these cases is that the pattern of pain often does not follow anatomical logic or neurological pathways. Gordon Waddell developed a series of clinical and bedside tests for “nonorganic” physical signs, which means symptoms and signs that do not follow physiological or anatomic logic (Table 16.4). These tests do not distinguish between malingering and sincere patients who are exaggerating, although they have been misused for this purpose. They demonstrate only that psychosocial factors are at play.

**Table 16.4.** Clinical Indicators of “Nonorganic” Low-Back Pain

Waddell Sign*	Inappropriate Response**
Superficial tenderness	Superficial sensitivity to touch; hyperesthesia
Non-anatomic tenderness	Tenderness to light touch that does not make sense anatomically
Axial loading	Vertical loading in a standing position: pushing down or putting weight on patient’s skull produces low-back pain.
Pain on simulated rotation	Passive, not active, rotation of shoulders and pelvis in same plane, with no flexion or extension, causes low-back pain. Spine does not rotate.
Distracted straight leg raise	Discrepancy between findings on sitting and supine straight leg raising tests; patients with true sciatica will find straight leg raising with flexion at the hip and straightening the leg while sitting equally uncomfortable. This test can be done unobtrusively while checking the Babinski reflex.

(Continued)

**Table 16.4.** (*Continued*)

---

Regional weakness	A weak muscle is easily overcome by force. This refers to weakness of muscle groups in an aberrant “cogwheel” pattern, in which initial resistance intermittently gives way.
Regional sensory change	Nondermatomal sensory loss; hypesthesia that does not follow dermatomes or nerve pathways, such as loss of sensation in an entire limb or one entire side of body
Overreaction	Disproportionate facial expression, verbalization, or tremor during examination, grimacing or other exaggerated response to pain that is not repeated when the same maneuver is performed later.
(Pain perception distorted) [Not an original Waddell sign]	Pain drawings made by patients also show non-physiological characteristics.

---

\*Using original terminology of Waddell (1980).

\*\*Three signs or more suggest a psychosocial component to disorder.

Most people with back pain manage the problem themselves or with the assistance of a chiropractor. Chiropractic manipulation is of unproven benefit beyond short-term palliation lasting a few hours. Even so, it is exceedingly popular with patients in the absence of definitive and effective medical management. Regardless of the legitimately critical attitude of mainstream medicine toward chiropractic, the public has clearly identified the system as a preferred mode of therapy for low-back pain. If spinal manipulation is performed gently and is limited to the lumbar spine, the risks of injury are low. The transient improvement that patients usually report may make it possible to return to normal life in comfort for

a period, and certainly reinforces the patient's use of the treatment. Chiropractic manipulation is most likely to be hazardous when it is forceful, applied to the cervical spine, and applied to patients who are ill with other conditions or of advanced age. Insurance companies tend to accept chiropractic management uncritically because it is cheap and placates the patient.

Many regimens for back rehabilitation have been proposed and few have shown evidence of appreciable benefit. This has led to a nihilistic attitude toward back interventions. Mainstays of the management of chronic back pain include modalities for the management of chronic pain (see the end of this chapter), graded exercise to increase conditioning and flexibility, limited use of analgesics on scheduled dosage for short periods, and intensive educational and rehabilitation programs. The most common approach to education is the organized "back school," in which patients are instructed in lifting techniques, exercises to strengthen their backs and the abdominal musculature that braces the trunk, and re-injury prevention. It is not clear whether these modalities are effective.

The indications for surgery are actually few in the context of occupational back pain. Surgery is often a misguided way out of an otherwise seemingly hopeless situation, but it risks making a bad situation much worse. Medical indications for surgical consultation are beyond the scope of this discussion, but they generally involve a clear indication of disc disease or other anatomical abnormality other than mild degenerative changes. Unfortunately, surgery has been heavily overutilized, with a generally poor success rate and occasional catastrophic exacerbation of the patient's problem. Re-operation is rarely helpful. There are, of course, exceptions and particular lesions that may require surgery in the individual patient, but these do not explain the heavy volume of spinal surgery being performed or the poor outcomes reported.

The factors that predict eventual recovery from chronic pain are social and psychological rather than physical or biomechanical. Successfully rehabilitated patients are more likely to have a guarantee of reemployment on recovery, and to have a longer tenure in the job

and higher performance ratings prior to the injury. They are less likely to be clinically depressed and to have ongoing litigation or appeals for compensation.

The degree to which delayed adjudication aggravates these cases is only now beginning to be fully appreciated. The psychological damage done by the continued uncertainty and the incentive to "hold on" to one's symptoms rather than to let them go is undoubtedly immense. Indeed, it has been suggested that "disability" from chronic low-back pain is an epidemic quite separate from the incidence of pain in the lower back. The position taken by Waddell and others is that social factors, psychological adjustment, overemphasis on the pain itself, and cultural expectations combine to create disability where tolerable discomfort would have otherwise been present. This is quite different from proposing that each patient is consciously exaggerating or exploiting a trivial complaint; this is clearly not the case. Rather, individuals in this society are subject to influences that tend to reinforce their sense of hopelessness and their emotional reaction to the pain. These factors contribute further by involving the patient in a lengthy adjudication process that forces obsession with the pain and constant introspective monitoring of subjective feelings that must reinforce the response. Moving patients through this is difficult but imperative; singling them out and pushing them back to work without psychological preparation is easy but does little more than add to their frustration.

The most common approach recommended in lifting heavy items is to squat and lift using extension of the legs with slow and deliberate movement, bringing the object through the separated knees, keeping the back straight, and holding the object close to the axis of the trunk. This recommendation is difficult or impossible to follow for many, perhaps most, lifting situations, however. Many items are simply too large to hold close to the body. For irregular or oversize objects, moving or animate loads, or items that may shift weight easily, usual recommended lifting procedures are largely impractical. In the absence of mechanical devices, slow and deliberate lifting is important because abrupt changes in position are most likely to lead

to acute injury. Bending and twisting motions are particularly risky because of the irregular dynamic stresses that they impose. The risk of injury also increases if the load is bulky and must be lifted from floor level.

A variety of other factors contribute to back injuries. Poor conditioning may result in lax muscle tone, lack of endurance required to sustain a lift safely, and poor flexibility. Short-term bursts of physical strength can be achieved even by relatively out-of-shape persons, so brute strength is not an indication of conditioning. Work habits that are rushed, inconsistent, pressured, and impatient often lead to errors in positioning or movement. Distractions and anxiety may substantially alter the perception of pain from back injuries and the effectiveness of rehabilitation. Dissatisfaction with work is a factor in some cases, not to be confused with overt malingering. Aging affects the back in many ways that may increase the risk of back injury, as do congenital or acquired back disorders that aggravate lordosis or that flatten the normal lordotic curve. Obesity is a major contributing factor, as abdominal weight places a directly shear force on the lumbar spine and also prevents holding loads close to the axial trunk. Vibration, as in vehicles or the operation of power tools, is a contributing factor for back pain in static positions, such as sitting for prolonged periods. Cigarette smokers also seem to be more at risk for disabling occupational back pain than are nonsmokers, possibly because of interference with perfusion to intervertebral discs.

Because of the formidable nature of the problem, employers have been attracted to measures that are intended to identify workers at risk for developing low-back pain. Personal characteristics such as age, sex, and anthropometry (body measurement) alone do not reliably predict whether a given worker will subsequently develop work-related back pain. Screening applicants for jobs as part of “pre-employment” screening programs (which are forbidden under the Americans with Disabilities Act, as opposed to the more modern pre-placement evaluations) was once popular among employers. A pre-employment radiographic viewing of the lumbar spine is a practice that is now considered both unethical, because of the radiation

hazard, and useless. In the absence of major and usually obvious deformities or past injury, lesions of the lower back that are visible on x-ray (such as lipping, minor wedging, and reduced disc space) have been convincingly shown to have little or no correlation with the subsequent development of work-related low-back pain. The amount of radiation involved is much too great to accept this unreliable test for screening purposes. In recent years, ultrasonographic measurement of the diameter of the spinal canal has been used to identify workers who have a spinal stenosis. This relatively subtle congenital anomaly seems to be related to an increased incidence of back pain, but it should not be used to exclude workers from the workplace because it is not an indicator of imminent risk to themselves or others, which is the only justification for fitness-for-duty screening that is not explicitly tied to work performance.

A deceptively promising approach to controlling occupational back pain in the new employee, in the context of pre-placement evaluations (see Chapters 15 and 18), was fitness-for-duty (also called fitness-to-work) evaluations, identifying workers with past histories of back pain or overt injury, and assessing their capacity to do the work in individual cases. The available screening tools are largely limited to a physical examination, stressing signs of back injury and deformity, and the conduct of a back fitness test, of which one of the most familiar is the National Back Fitness Test. These tests have not been well documented or standardized, and some, such as the National Back Fitness Test, have been criticized for incorporating elements that may be hazardous (in this case, the double straight-leg raise maneuver). In the future, fitness-to-work evaluations involving the back are likely to be facilitated by mechanical screening techniques now in development. These may include static or isometric strength testing, dynamic strength testing, and endurance testing. The logic behind these programs is suspect. These are tests of work capacity, not of risk of back pain. The ability to lift, twist, and hold a position at one point in time has little relevance to risk of back pain at another point in time; and even if this were not the case, such a test would be unacceptable because it might precipitate low-back pain.

A better approach is to evaluate the job and its individual motions and to make reasonable efforts to prevent unnecessary stress on the back.

Occupational low-back pain is preventable by a combination of approaches as outlined in Table 16.5. These approaches eliminate the largely discredited approach of screening new applicants, and instead emphasize fitness-to-work evaluation, ergonomic design, fitness, training and education, and efforts to improve working conditions and attitudes. When back injuries do occur, their recurrence and their severity can be minimized by appropriate management with activity, early rehabilitation and return to work, and education to prevent re-injury.

**Table 16.5.** Elements of a Back Pain Prevention Program

---

1. Job Analysis and Redesign
  - 1.1 Identify risk factors (load, movements, vibration, prolonged sitting and static postures, awkward lifts)
  - 1.2 Material handling equipment
  - 1.3 Stools, benches, other furniture
  - 1.4 Ergonomic design of workplace
  - 1.5 Reduce loading requirements (e.g., smaller units)
  - 1.6 Job rating and detailed description for use in fitness-to-work assessment
2. Assessment of Fitness to Work
  - 2.1 Medical screening for history of back pain (not to include x-ray screening)
  - 2.2 Strength and fitness testing
  - 2.3 Work-hardening and rehabilitation for injured workers
3. Training
  - 3.1 Safe lifting practices
  - 3.2 “Back school”
4. Case Management
  - 4.1 Appropriate early treatment; minimal bed rest
  - 4.2 Early return to work
  - 4.3 Early settlement of compensation claims

*(Continued)*

**Table 16.5.** (*Continued*)

- 
- |                                  |  |
|----------------------------------|--|
| 4.4                              | Early intervention in depression   |
| 4.5                              | Vocational rehabilitation if return to work cannot be achieved                         |
| <br>5. Communication and Liaison |  |
| 5.1                              | Senior management, workers, supervisors, union representatives to be educated          |
| 5.2                              | Objectives of program explained and discussed; motives clear                           |
| 5.3                              | Evaluation and regular reporting of injury rates and performance of program as a whole |
| 5.4                              | Communication with workers' compensation carrier                                       |
- 

## REPETITIVE STRAIN INJURY

Repetitive strain injury (RSI) is a class of musculoskeletal disorders in which chronic discomfort, pain, and functional impairment may result from numerous repeated movements of the upper extremity. Swelling is often present. There are many synonyms for repetitive strain injury, including "overuse syndrome" and "cumulative trauma disorder." Although there are several specific conditions within the general category, all have in common the association of repetitive movement, rather than a single initiating event, and a subjective nature of the presenting complaints that makes diagnosis difficult and claims for compensation problematical. Women are more commonly affected, in large part because of the nature of the jobs associated with a high risk. The APG provides guidance in diagnosis and management.

The frequency of reported repetitive strain injuries has increased dramatically in recent years, led in the 1980s by Australia, where it emerged as the leading issue in occupational health. In Japan and the United States, as well, there was a dramatic increase in claims for RSI in the 1980s and 1990s. This may be due to increased recognition and attention to the disorder, but the strong impression of many occupational health authorities infrequency is that the increase in frequency is real. Most observers believe it is associated with the increasing number of jobs in the economy that relate to paced or work-driven

execution of a limited number of relatively fine motor movements of the hands and arms, such as keystrokes, assembling small parts, cutting fabric, and packaging small items. In the past, many RSIs were probably treated as isolated or exceptional cases without a conceptual framework for understanding and recognizing the broader problem. However, jobs requiring such repetitive actions have obviously proliferated in recent years, and workloads and work pacing have increased with the use of the Internet and electronic records. Occupations with particularly high rates of RSI have included airline reservation agents, telephone operators, insurance claim and other data input workers, stockbrokers, secretaries, and newspaper reporters, although some less automated industrial jobs such as meat packaging and textile manufacturers have also been greatly affected.

Occupational factors associated with RSI include a sustained and awkward posture, excessive manual force, use of intrinsically weak body parts in unusual or forceful movements, inadequate time for recovery or breaks in the work, and high rates of repetition of the movement. Load factors are important in producing RSI, and additional occupational hazards, specifically cold working conditions and vibration, may make the condition worse. The psychological effects of the stress and noise frequently associated with these jobs may also contribute to the subjective perception of pain and discomfort.

RSI is distinguished from non-occupational rheumatologic conditions, temporary pain from sprains or muscle strain, psychological disorders (including compensation neurosis), abnormalities of bony structure, and single-event injuries. In the United Kingdom, the Industrial Survey Unit of the Arthritis and Rheumatism Council concluded that persistent or recurrent musculoskeletal pain without immediate traumatic cause within the previous six weeks suggests the diagnosis. RSI is often considered to develop through three stages:

- i. Stage 1 is a condition of fatigue characterized by aching and tiredness increasing during the work shift, and is usually reversible with overnight rest.

2. Stage 2 is persistence of the discomfort into the next day and earlier onset of fatigue during the workday.
3. Stage 3 is chronic aching, fatigue, and weakness that persist despite rest of the affected part.

These stages are, of course, generalizations but they are very useful in explaining the progression of these conditions to patients.

Each of the specific disorders to be discussed has its own natural history. Some generalizations are possible, however. Because the symptoms are mostly subjective and gradual in onset, it can be difficult to obtain a satisfactory history for these conditions, particularly for onset. Initial treatment is usually conservative, which in this context involves rest, initial cooling with ice followed later by local heat, and anti-inflammatory medications. Splinting may be helpful for short periods, especially at night when spasms and pain disturb sleep.

More effective than treatment in controlling the problem is prevention by changing the nature of the operation, tools, or layout of the workplace to modify or eliminate the offending action. In every case, RSI is much easier to prevent than to treat. Effective prevention requires attention to the actual movements conducted by the worker in a given workplace and the ergonomic redesign of the job to minimize the hazardous movement patterns. This is often easy to do by rearranging the workstation, changing tools, or changing the pace of the work. Unfortunately, these measures are too often considered only after a number of cases have been identified and a relatively serious problem is recognized (Table 16.6).

Chronic musculoskeletal disorders are not difficult to diagnosis but can be hard to explain. After all, the human body regenerates itself and should not wear out like a machine part. This has led some observers to question the existence of musculoskeletal disorders. Norton Hadler, a rhetorically gifted iconoclast, is particularly vocal at that extreme of the spectrum of opinion. Such criticisms of the field as a whole typically focus too narrowly on carpal tunnel syndrome and the many known risk factors that are not occupational. They do not directly engage the

**Table 16.6.** Repetitive Strain Injury: Common Conditions and Diagnostic Features

Condition	Diagnostic Features
<i>Neck and shoulder girdle</i>	
Tension neck syndrome	None
Cervical syndrome	Pain with flexion/extension of neck radiating down arm
Thoracic outlet syndrome	Variation in pulse strength with hyperextension at shoulder (Adson's maneuver)
<i>Shoulder and upper extremity</i>	
Tendonitis, tenosynovitis, bursitis	Local pain, swelling
Supraspinatus tendonitis (rotator cuff tendonitis)	Pain on abducting beyond 70° at shoulder
Bicipital tendonitis	Pain over bicipital tendon
Frozen shoulder syndrome	Reduced range of motion after injury
Acromioclavicular syndrome	Pain over acromioclavicular joint when clavicle is percussed while patient is pushing downward against resistance
<i>Upper extremity and hand, wrist</i>	
Lateral or medial epicondylitis	Local pain and pain with resisted hand motion
de Quervain's tenosynovitis	Finkelstein's test
Carpal tunnel syndrome	Nerve conduction studies, Tinel's sign, Phalen's test
Ulnar nerve entrapment	Nerve conduction studies, Tinel's sign

huge body of literature that documents occupational factors among the many known risk factors for carpal tunnel syndrome and other upper extremity RSIs. Many of the studies cited as contrary evidence are cross-sectional in design and simply identify personal risk factors for carpal tunnel syndrome, which is known to have many nonoccupational causes. Cross-sectional studies do not adequately assess the preva-

lence of risk factors when workers have left the job because of early pain or difficulty managing on the job or because of active claims. The mere fact that personal risk factors may exist for many cases does not mean that occupational factors play no role for the large excess of cases seen in certain occupations. One would have to make a convincing case that occupational factors are never involved in RSI cases in order to deny the association; the weight of evidence, however, strongly suggests otherwise. RSIs may or may not be over-diagnosed, but they do exist.

An RSI (most often carpal tunnel syndrome in this context) may not have occurred “but for” the occupational factors in a person who is at risk due to personal risk factors that are necessary but not sufficient; this is probably a common situation. Workers’ compensation policies usually reject such claims and they go to appeal, in which case they may be deemed an “exacerbation” of preexisting disease.

### **Tension Neck Syndrome (Tension Myalgia)**

This common condition is characterized by a persistently stiff, aching neck often accompanied by a headache. It is the result of static, sustained muscle contraction, accompanied by local spasms of the trapezius and other neck muscles, and is considered by many to be a form of fibrositis (an ill-defined condition that may not be an actual disease) because of the pressure of nodules or trigger points. It is reported to occur frequently among typists, keypunch operators, cashiers, small parts assembly workers, packers, and others who must maintain a restricted posture with activity of their forearms while bracing at the shoulder. Tension neck syndrome is treated conservatively. Relaxation exercises, application of heat, fitness training, and a soft cervical collar may help. Consciously altering posture, being aware that the chin is kept tucked in, and stretching exercises frequently help.

### **Cervical Syndrome**

This condition resembles the tension neck syndrome in some patients but also involves pain radiating to one or both arms, and may

be associated with numbness or paresthesias in the hands. Range of motion of the neck is usually restricted by pain and there may be reduced power in the deltoid, triceps, and biceps, although this may be difficult to document. Cervical syndrome is not infrequent among individuals who must repeatedly flex or hyperextend their necks and assume awkward positions for long periods, including cash register operators, data entry typists, painters, decorators, and dental surgeons. The process appears to be one of disc degeneration with age, with aggravation of the weakened vertebral joint from exaggerated positions. Overt cervical osteoarthritis may be present radiographically, but patients may complain of all the symptoms without bony abnormalities. Bony changes can often be seen on neck films, including spondylosis or osteophytes. On the other hand, osteophytes are not uncommon in individuals with no symptoms, as is evident when a cervical spine film is obtained for other reasons. Conservative treatment, including ice or heat, muscle relaxants, and massage may provide marked relief, as may ultrasound. A cervical collar, initially soft and then progressing to hard if necessary, can provide substantial symptomatic relief.

### **Thoracic Outlet Syndrome**

A neurovascular condition, thoracic outlet syndrome occurs when the nerves of the brachial plexus and the brachial artery and its branches are compressed between the muscles of the neck and shoulder. The result is numbness, ischemia, and pain in the distal upper extremity, especially in postures in which the shoulders are thrown back and the hand is raised. A clinical test for this (the Adson maneuver, Figure 16.1) is to hyperextend the shoulder, in the manner of a military order to “stand at attention,” with the chin thrust forward. A weakened pulse and reproduction of the symptoms strongly suggest the thoracic outlet syndrome. The symptoms of the syndrome may also occur in the presence of atherosclerosis of the brachial artery, in which case the test may not be effective (even though pain, numbness, and weakness may appear during work with the arms). It may



**Figure 16.1.** Adson's maneuver for thoracic outlet syndrome. Reprinted from "Occupational Repetitive Strain Injury," Issue 45, Volume 2: 585-92, *American Family Physician*. Copyright © 1992 American Academy of Family Physicians. All Rights Reserved.

also occur in individuals with congenital cervical ribs or abnormal muscle placement and insertions. It is usually found in occupations requiring frequent reaching above shoulder level, prolonged carrying of relatively heavy loads (such as suitcases) at the side of the body, wearing a knapsack or other item with straps around the shoulder, or bracing with the shoulders while carrying a stretcher or similar load in a fixed position at waist level. Occupations at risk include grinders, overhead assembly workers, auto repair mechanics, cashiers, musicians, operating

room personnel, truck drivers, stockroom and shipping workers, and letter carriers. In advanced or severe cases, surgery may be required, but often merely avoiding the offending motion is sufficient.

### **Tendonitis, Tenosynovitis, Bursitis**

Tendons are lined by a synovial sheath (except in the shoulder) and are cushioned at points of stress and leverage by bursae. Inflammatory conditions of these structures are very common, and it is important to rule out septic tendonitis or tenosynovitis in any case in which repetitive movement or a traumatic cause is not obvious. RSIs can affect virtually any mobile part of the body given sufficient repetitive loaded motion. For mechanical and loading reasons, however, they are most common in the upper extremity, although they are not infrequently seen in the lower extremity among athletes. Tendonitis results when the muscle-tendon connection is repeatedly tensed and the tendon begins to fray or tear apart. Unless allowed to recover fully, the tendon may become weakened through repeated re-injury. The tendon sheath may develop an effusion and inflammation (tenosynovitis), particularly in the hand. Where the tendon lacks a sheath, as in the shoulder, the injured tendon may calcify. When the sheath becomes thickened, it may constrict movement of the tendon, causing a stenosing tenosynovitis such as de Quervain's disease (see below). Another common example is "trigger finger" (stenosing tenosynovitis crepitans), in which attempts to flex a finger are impeded; the tendon "gives" abruptly in a jerking movement, like a knotted string being pulled through a hole in a piece of cardboard. These conditions are most common in the hand, where the muscle and tendons are mechanically weak, but also occur at the elbow and shoulder (and contribute to "tennis elbow" and related conditions). These conditions are common whenever there is repetitive motion on the job, and are particularly problematic when hand tools are used. Such conditions are frequently encountered among grinders, machine tool operators, assemblers, sewers and cutters, musicians, packers, and meat packers. Treatment

is conservative with physiotherapy for range of motion exercises. Braces or splints may be useful depending on the part affected. Local injection of steroids should be minimized to avoid weakening the tendon permanently.

### **Supraspinatus Tendonitis (Rotator Cuff Tendonitis)**

This condition is an inflammation of the tendon of the supraspinatus muscle, the muscle involved in initially abducting the humerus from neutral position under the acromion, and in an elevated position pushes against the acromion painfully when inflamed. It is common among workers who must maintain a position of shoulder abduction with the elbow extended under conditions of load, including welders, painters, aluminum siding and awning installers, riveters, and construction workers. Characteristic pain with abduction between 70° and 100° is the principal diagnostic feature, called the “impingement syndrome.” Treatment is often unsatisfying because the condition waxes and wanes and frequently returns with re-use. Rest, heat, anti-inflammatory medications, and physiotherapy can help the condition, and it is important to maintain range of motion exercises in order to avoid developing frozen shoulder syndrome. Local injection of steroids is used by some but has not been convincingly shown to be of benefit.

### **Bicipital Tendonitis**

Bicipital tendonitis often occurs in association with supraspinatus tendonitis, but may also occur in isolation. This condition is similar to supraspinatus tendonitis, involving pain with movement of the glenohumeral joint and pain over the bicipital tendon as it passes over the bicipital groove and under the acromion. The condition is very common among workers who must reach up over their heads, such as assembly workers, cleaners and window washers, construction workers, and stockroom and shipping clerks. Treatment is similar to that for supraspinatus tendonitis.

### **Frozen Shoulder Syndrome (Adhesive Capsulitis)**

This is a condition of contracture in the soft tissues surrounding the glenohumeral joint. Thickening of the tendons and bursae result in fixing, and in severe pain with freed motion. The condition typically results from prolonged immobilization after shoulder injury and is a complication common to many soft tissue injuries involving the shoulder. It presents as progressive chronic pain and stiffness developing over several months, and results in loss of active and passive range of motion at the shoulder. It is difficult to treat but can be prevented by physiotherapy and range of motion exercises during recovery from shoulder injuries. Frozen shoulder can also occur spontaneously, and is most common in middle-aged women.

### **Acromioclavicular Syndrome**

This is a condition similar to the forms of tendonitis mentioned above but distinguished by local pain over the acromioclavicular joint. Its presence can be confirmed by having the patient push downward against resistance and percussing the clavicle; the pain should be reproduced. This syndrome occurs as a result of repeated movement with loaded stress on the joint at waist level, as might occur in grinding, packing, assembly, and construction work. Treatment is conservative, but surgery is sometimes required in extreme cases; local injection of steroids may help.

### **Epicondylitis**

Two forms of epicondylitis are commonly named for their counterpart sports injuries: “tennis elbow” (lateral epicondylitis) and “golfer’s elbow” (medial epicondylitis). They are both inflammatory conditions of the tendonous origins of the finger muscles involving the extensors and the flexors, respectively. Both arise from repeated and forceful rotation of the forearm with the wrist bent, the different locations reflecting different patterns of stress at the elbow. Pain over the lateral epicondyle with palpation and during extension of

the fingers and waist against resistance with the elbow straight suggests lateral epicondylitis. Pain over the medial epicondyle and with flexion of the fingers and wrist against resistance with the elbow flexed suggests medial epicondylitis. The conditions are common in jobs requiring repeated arm and elbow motions as described: small parts assemblers, musicians, construction, and woodworkers. Treated conservatively and with local injection and range of motion exercises, these conditions often take months to resolve.

### **de Quervain's Tenosynovitis**

The most common stenosing tenosynovitis, de Quervain's disorder, involves the extensor tendons of the thumb and is most characteristic of workers using hand tools in repeated motion. The thumb tendon (extensor pollicis brevis) is affected by nodules in the tendon surface slipping through the fibrosed parts of the tendon sheath. Finkelstein's test, in which the patient clenches his or her fist over the flexed thumb and the examiner forcefully pushes the base of the thumb toward the ulna, is diagnostic when there is pain at the radial styloid process (see Figure 16.2). Management is usually conservative with



**Figure 16.2.** Finkelstein's test for de Quervain's tenosynovitis. Reprinted from "Occupational Repetitive Strain Injury," Issue 45, Volume 2: 585-92, *American Family Physician*. Copyright © 1992 American Academy of Family Physicians. All Rights Reserved.

steroid injection, although surgical correction is possible and effective. The evidence for an occupational association with repetitive strain is weaker with de Quervain's tenosynovitis than for other musculoskeletal disorders.

### **Traumatic Mononeuropathies**

These conditions are injuries to peripheral nerves occurring as a result of direct trauma or repetitive strain injury. Obviously, injuries to any part of the body may cause local injury directly. Recovery from injury or the effects of repetitive motion on soft tissue surrounding the nerve may cause entrapment of the nerve.

Carpal tunnel syndrome is discussed in detail below with respect to repetitive strain injury, the most common cause of this nerve entrapment syndrome affecting the median nerve. Radial and ulnar entrapment neuropathies are much less common.

Pressure neuropathies occur when nerves in exposed or vulnerable positions are subjected to pressure. Some of the most common of these are temporary paralysis of the radial nerve (by anything compressing the arm above the elbow), the posterior interosseous nerve (by bricks placed on the forearm: "bricklayer's palsy"), and the axillary nerve (by anything held in the armpit). Other common neuropathies include the lateral femoral cutaneous nerve (by tools, long rods, or poles supported against the anterior superior iliac crest), the peroneal nerve (by tight boots or injury to the leg), and the sciatic nerve (from prolonged sitting, especially on horseback).

### ***Carpal Tunnel Syndrome***

In this nerve entrapment condition, the median nerve is compressed along its path in the narrow channel in the wrist defined by the radius, the flexor retinaculum, and the tendons of the muscles for flexion of the hand. This causes paresthesias and numbness in the distribution of the nerve, including the volar aspects of the first four

digits and the thenar area of the palm, and the distal dorsal aspects of digits II, III, and (partially) IV. If it progresses untreated, atrophy of the thenar muscles may also occur and fine motor movements may become difficult to execute. The pain and tingling often disturbs sleep. The condition is usually associated with repeated forced hand movements such as those that occur among cashiers, assembly workers, grinders, typists, keypunch operators, sewers and cutters, musicians, packers, and bricklayers. Diagnosis is not usually difficult; this is one of the few occupational RSIs to have become widely recognized. Carpal tunnel syndrome is often associated with a number of other repetitive strain injuries in the same patient. It is also a complication of a number of medical conditions and of pregnancy. Tingling paresthesias with percussion over the median nerve at the wrist (Tinel's sign) is strongly suggestive and can be confirmed by nerve conduction studies showing slowed nerve impulses, although this tends to occur relatively late, when the entrapment is far along. Tinel's sign has relatively low sensitivity and specificity, and is positive in any partial nerve damage or hyper-irritability. A relatively specific, but not very sensitive, clinical test is Phalen's maneuver, in which the wrists are passively flexed by press-



**Figure 16.3.** Phalen's test for carpal tunnel syndrome. Reprinted from "Occupational Repetitive Strain Injury," Issue 45, Volume 2: 585-92, *American Family Physician*. Copyright © 1992 American Academy of Family Physicians. All Rights Reserved.

ing the flexed hands together dorsum-to-dorsum for one minute in an effort to elicit the pain and paresthesias (see Figure 16.3). Conservative treatment, including a splint, often fails to provide relief; surgery to release the entrapped nerve is usually effective.

### ***Ulnar Nerve Entrapment***

The ulnar nerve passes behind the medial epicondyle and the elbow, and is in a position vulnerable to repeated minor trauma. The forearm may be rested on a hard chair arm or workbench. In time, the injured nerve may become entrapped by local swelling and tissue hypertrophy, thus becoming even more susceptible to minor trauma of this type. Preceding signs are similar to those for carpal tunnel syndrome except that the paresthesias are felt over the ulnar side of the hand and over digit V. Diagnosis is similar to that for carpal tunnel except that Tinel's sign can be demonstrated over the epicondyle. Treatment is also similar, except that there is no surgical intervention.

## **CHRONIC PAIN**

Many occupational injuries involve pain, and the evaluation and treatment of chronic and repetitive musculoskeletal disorders, particularly soft-tissue injuries, is dominated by issues of management and interpretation of chronic pain. The OEM physician is guided in this complex area by the *APG*, specifically the module Chronic Pain (the current edition at the time of writing was 2008), which should be considered authoritative.

Acute pain is the familiar mechanism of pain perception in response to acute stimulation. Acute pain is best managed by treating the underlying cause of the pain and by analgesia appropriate to the severity of the pain. Acute pain is a signal that something is wrong and, during the early days of recovery, it helps in avoiding further injury. "Chronic pain" is qualitatively different. During recovery,

some injured workers develop a state in which pain becomes dysfunctional, or the degree of pain does not match the functional impairment. Their depressed mood reflects the perception of pain, and the persistence of pain becomes an organizing principle in life. Chronic pain overwhelms recovery to become the central problem and looms to the patient (because the injured worker has psychologically “regressed” to a sick role) as an insurmountable obstacle to any return to normal life. Chronic pain is therefore a condition that is best managed by what is termed the “biopsychosocial model,” taking into account emotional and perceptual factors, rather than by the medical model, which involves searching for structural lesions and specific functional impairment. There is evidence to suggest that many physical modalities (such as passive stretching or manipulation) may actually make chronic pain worse but that activity (such as aerobic exercise) may improve outcomes.

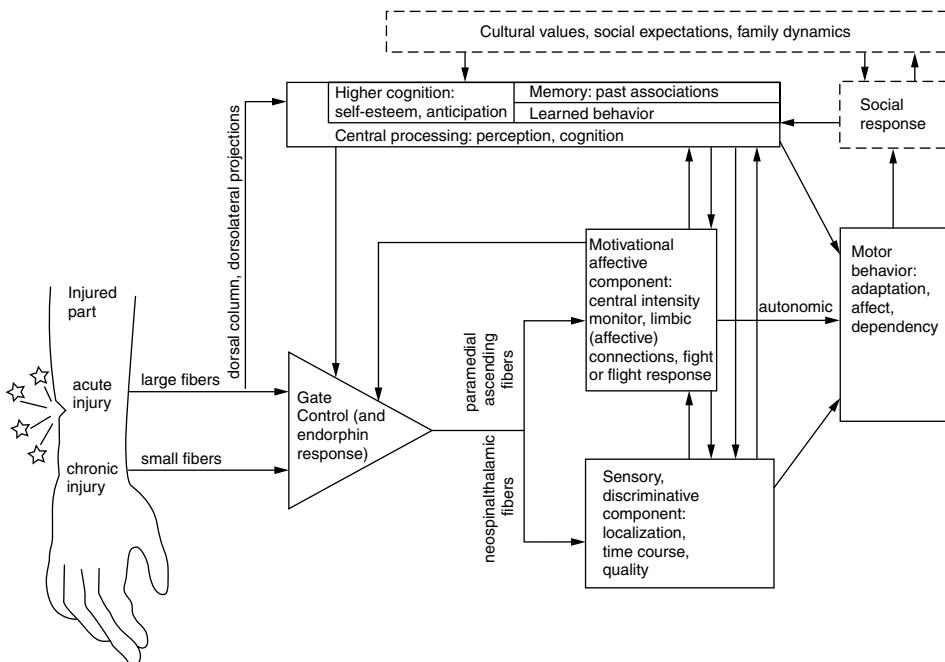
Pain is a complex sensation, only partly the result of stimulation of particular pain receptors and transmission of nerve impulses to the brain. Pain is integrally linked with affective and cognitive functions, and modulated by neural and neurohormonal influences in a complex manner. Virtually all of the complexities of pain perception come into play in occupational injuries because of the social dynamics of occupation and compensation.

Somatic pain stimulation varies in intensity depending on the extent of injury, the nature of the injury, the site of injury, and the extent of use of the body part. The impulse is carried by afferent pathways to the brain by way of the spinal cord, where a proposed mechanism called “the gate” modulates transmission of pain impulses. This gate acts as a control device to dampen the intensity of perceived pain or to increase its perception, depending on whether the transmission of pain impulses is impeded or facilitated. The gate tends to “close,” or impede transmission, when other sensations accompany the pain, and tends to “open,” or facilitate transmission, when acted on by higher cortical influences. Endogenous endorphin release also appears to be an important

mechanism in damping the perception of pain, although this mechanism is not part of the familiar “gate control theory.” Cognitive states, such as anxiety and rumination, and emotions, such as depression, dramatically affect the perception of pain, as do poor sleep quality, reduced competing sensory input as at night when the injured worker is trying to sleep, and drawing of attention to it by others. Distraction, on the other hand, can reduce the perception of pain equally dramatically. This is a simplistic description of a very complicated mechanism.

Once the pain sensation reaches the brain, there are additional cognitive and affective mechanisms that associate pain with avoidance behavior (in addition to spinal reflexes), emotional responses, and thought processes related to interpretation of the significance of the pain. Chronic pain has a strong cortical component and is characterized by an affective response to the pain that is out of proportion to the physical injury, and inappropriate for the point of time in recovery. Figure 16.4 presents a schema (adopted from Melzack) integrating the various elements of pain perception and response. It appears that acute pain and chronic pain are conducted by different pathways. The pathway for acute pain (the “lateral pathway”) is rapid, localized, and phasic; the pathway for chronic pain (the “medial pathway”) is slow, diffuse, persistent and deeply anxiety provoking. These differences may explain why acute injuries seem to be specific and why throbbing is common (although vascular sensations are also important), while chronic pain often seems to be generalized, relatively constant, and depressing. Pain theory has advanced since Melzack’s model, but the schema is a robust, easily understood framework that provides useful guidance.

It is obvious that the perception of pain may affect mood. Additionally, mood and cognition may affect the perception of pain. Clinically, a depressed mood, a lack of distracting interests, and cognitive focus on the injury are associated with increased perception of pain. Mood elevation, active physiotherapy and other activities, involvement in the community, and absorption with intellectually



**Figure 16.4.** Conceptual model of pain, as it applies to occupational injuries. Reprinted from “Occupational Repetitive Strain Injury,” Issue 45, Volume 2: 585–92, *American Family Physician*. Copyright © 1992 American Academy of Family Physicians. All Rights Reserved.

stimulating and rewarding activities all seem to reduce the perception of pain. Part of the cognitive intervention in chronic pain cases is to reassure the patient that the pain that he or she experiences is not a threat to health or evidence of ongoing injury and does not have the same meaning as acute pain.

Unfortunately, occupational injuries by their nature involve documentation, claims, often appeals, sometimes legal action (against third parties), and endless repetition of the story of the incident to examining physicians, adjudicators, case managers, family, neighbors. Perversely, the emphasis in claiming benefits is often on how badly the injured worker is impaired, not on how well he or she is recovering. The result is that the injured worker is constantly reliving the incident, and thus it assumes a defining role in his or her life. Compensation claims are often disputed and must be appealed, physicians must be consulted,

treatment must be scheduled, the worker's life is disrupted, and, at the extreme, family and friends may question whether the hapless patient is exaggerating to gain sympathy or other benefit. The result is that the injury may become the center of the injured worker's life, a source of great frustration to which are connected issues of self-esteem, community recognition, relationships with family and others, fulfillment of goals, blaming of others, and control over one's own life. The management of chronic pain in such situations should be integrated, using medical and psychotherapeutic approaches. Realistically, there is little hope for full resolution until the status of the claim is resolved. Dependency is a serious risk during this period.

Increasingly, pain management centers are becoming available to serve as resources for patients with chronic pain from various disorders. These centers emphasize the development of strategies for coping with pain by minimizing its effects on life, avoiding brooding and emotional identification with the pain. Long-term analgesia is discouraged in favor of psychological approaches that promote independence. Chapter 15 addresses the issue of somatization and pain perception.

A concomitant problem in chronic pain is analgesic dependence and substance abuse. In an attempt to be totally pain free, the injured worker may insist on opiates, which should be given only as appropriate for acute pain, or may turn to other pharmacological means of reducing pain and feeling better. This chemical shortcut is difficult to give up once the patient has started down that route, because he or she is tempted to resort to immediate relief rather than allowing longer-term interventions the necessary time to work. Management of chronic pain by analgesics is not recommended; however, attempts to so manage have been highly problematical even among specialized chronic pain treatment centers and physicians who devote their practice to this area. The OEM physician who undertakes treating chronic pain patients is advised to obtain special training and preparation. Biological and behavioral means of altering pain sensation generally work better for chronic pain than do physical interventions.

There are special chronic pain syndromes, each of which requires special management beyond the scope of this chapter:

- “Complex regional pain syndrome,” previously called reflex sympathetic dystrophy, is an uncommon condition in which a body part develops pain of a burning, neuropathic quality, often with sympathetic symptoms such as vasodilation, swelling, sweating, hyperesthesia, and sometimes atrophy or severe weakness. This syndrome is not well understood and rarely occurs spontaneously.
- “Causalgia” (often called “Type II complex regional pain syndrome,” and historically called “hot pain”) is an intense neuropathic pain, otherwise similar to the complex regional pain syndrome that follows healing from a specific severe nerve injury, such as a gunshot wound or crush injury.
- “Neuropathic pain,” involves damage to nerves and, presumably, incomplete recovery. The injury may be toxic, metabolic (as in diabetic neuropathy), or traumatic. A characteristic of neuropathic pain is that it has a relentless burning quality and can be highly refractory. Neuropathic pain in general is often more responsive to antidepressants (amitriptyline is usually used), anticonvulsants, and serotonin-norepinephrine reuptake inhibitors than to nonopiate analgesics.
- “Fibromyalgia” is a complicated syndrome, which some would call a disease, consisting of muscle pain and well-defined trigger points.
- “Myofascial pain” and isolated “trigger points” are regional conditions with characteristics similar to fibromyalgia.

## RESOURCES

- American College of Occupational and Environmental Medicine. *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery in Workers*. 2nd ed. (rev). Chicago: American College of Occupational and Environmental Medicine; 2008.
- Derebery J, Anderson JR. *Low Back Pain: An Evidence-Based, Biopsychosocial Model for Clinical Management*. 2nd ed. Beverley Farms, MA: OEM Press; 2008.

## Occupational Injuries and Musculoskeletal Disorders

- Haig AJ, Colwell M. *Back Pain*. Philadelphia: American College of Physicians; 2005.
- McCarberg B, Passik SD. *Expert Guide for Pain Management*. Philadelphia: American College of Physicians; 2005.
- National Institute for Occupational Health and Safety (NIOSH). *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-Related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back*. Publication 97-141. Cincinnati: NIOSH; 1997.

### NOTEWORTHY READINGS

- Australia National Health and Safety Commission. *Repetition Sprain Injury: A Report and Model Code of Practice*. Australian Government Publishing Service: May 1986.
- Ayoub MA. Ergonomic deficiencies: I. pain at work. *J Occup Med*. 1990;32:52–57.
- Bergenudd H, Nilsson B. Back pain in middle age; occupational workload and psychologic factors: an epidemiologic survey. *Spine*. 1988;13:58–60.
- Bergenudd H, Nilsson B. The prevalence of locomotor complaints in middle age and their relationship to health and socioeconomic factors. *Clin Orthop*. 1994;308:264–270.
- Bigos SJ, Battie MC. Preplacement worker testing and selection considerations. *Ergonomics*. 1987;30:249–251.
- Bigos SJ, et al: The prospective study of work perceptions and psychosocial factors affecting the report of back injury. *Spine*. 1991;16:1.
- Binder AI, Hazelman BL. Lateral humeral epicondylitis—a study of natural history and the effect of conservative therapy. *Br J Rheum*. 1983;22:73–76.
- Bleeker ML, Agnew J. New techniques for the diagnosis of carpal tunnel syndrome. *Scand J Work Environ Health*. 1987;13:385–388.
- Boden LI, Ozonoff A. Reporting workers' compensation injuries in California: how many are missed? Sacramento, Labor and Workforce Development Agency, Department of Industrial Relations. August 2008. <http://74.125.95.132/search?q=cache:nu9canRVyFYJ:www.dir.ca.gov/chswc/reports/Reporting%20WorkersCompensationInjuriesinCalifornia2008August.pdf+California+work+lost-time+injuries+statistics&hl=en&ct=clnk&cd=6&gl=us>. Accessed January 6, 2009.
- Chan CW, Goldman S, Illstrup DM, Kunselman AR, O'Neill PI. The pain drawing and Waddell's nonorganic physical signs in chronic low-back pain. *Spine*. 1993;18(13):1717–1722.
- Cyriax JH. The pathology and treatment of tennis elbow. *J Bone Joint Surg*. 1936;18:921–940.
- Dagenais S, Caro J, Haldeman S. A systematic review of low back pain illness cost in the United States. *Spine*. 2008;8(1):8–20.
- Denault B. How to follow up a workplace accident. *OH&S Canada*. 1990;6(2):86–96.
- Deyo RA, Diehl AK, Rosenthal M. How many days of bed rest for acute low back pain? *N Engl J Med*. 1986;315:1064–1070.
- Feldman RG, Goldman R, Keyserling WM. Peripheral nerve entrapment syndromes and ergonomic factors. *Am J Ind Med*. 1983;4:661–681.

## The Praeger Handbook of Occupational and Environmental Medicine

- Gallagher RM, Rauh V, Haugh LD, Milhous R, Callas PW, Langelier R, McClallen JM, Frymoyer J. Determinants of return-to-work among low back pain patients. *Pain*. 1989;39:55-67.
- Gibson ES, Martin RH, Terry CW. Incidence of low back pain and pre-placement x-ray screening. *J Occup Med*. 1980;22:515-519.
- Harris JS, Brigham C. Low back pain: impact, causes, work relatedness, diagnosis, and therapy. *OEM Rep*. 1990;4(11-12):84-88, 92-96.
- Inancsi W, Guidotti, T.L. Occupation-related burns: five-year experience of an urban burn center. *J Occup Med*. 1990;29:730-733.
- Kasdan ML, ed. *Occupational Hand and Upper Extremity Injuries and Diseases*. Philadelphia: Hanley and Belfus; 1990.
- Katz JN, Larson MG, Fossel AH, Liang MH. Validation of a surveillance case definition of carpal tunnel syndrome. *Am J Pub Health*. 1991;81:189-193.
- Katz JN, Larson MG, Sabra A, Krarup C, Stirrat CR, Sethi R, Eaton HM, Fossel AH, Liang MH. The carpal tunnel syndrome: diagnostic utility of the history and physical examination findings. *Ann Int Med*. 1990;112:321-327.
- Kern D, Zlatkin MB, Dalinka MK. Occupational and post-traumatic arthritis. *Radiol Clin N Am*. 1988;26:1349-1358.
- Melzack R. Physiology of pain. *Med N Am*. 1989;3(36):6524-6531.
- Nachemson A, Jonsson E, eds. *Neck and Back Pain: The Scientific Evidence of Causes, Diagnosis, and Treatment*. Philadelphia: Lippincott Williams & Wilkins; 2002.
- Phalen GS. The carpal tunnel syndrome, clinical evaluation of 598 hands. *Clin Orthop Rel Res*. 1972;83:29-40. [A classic.]
- Pick RY. De Quervain's disease: a clinical triad. *Clin Orthop Rel Res*. 1979;143:165-166.
- Polatin PB, Gatchel RJ, Barnes D, Mayer H, Arens C, Mayer TG. A psychosocial-medical prediction model of response to treatment by chronically disabled workers with low-back pain. *Spine*. 1989;14:956-964.
- Putz-Anderson V, ed. *Cumulative Trauma Disorders: A Manual for Musculoskeletal Diseases of the Upper Limbs*. New York: Taylor and Francis; 1988.
- Putz-Anderson V, Pizatella T, Tanaka S. *A Proposed National Strategy for the Prevention of Musculoskeletal Injuries*. National Institute for Occupational Safety and Health, and Associations of Schools of Public Health. 1986:17-34. [NTIS 87114740]
- Quebec Task Force on Spinal Disorders. Scientific approach to the assessment and management of activity-related spinal disorders. *Spine*. 1987;12(suppl 1):7S.
- Spengler DM, Bigos SJ, Martin NA, Zeh J, Fisher L, Nechmemson A. Back injuries in industry: a retrospective study. *Spine*. 1986;11:241-245.
- Tichauer ER, Gage H. Ergonomic principles basic to hand tool design. *Am Ind Hyg Assoc J*. 1977;38:622-634.
- Trow GWS. A reporter at large: devastation. *New Yorker*. October 22, 1990;nv:54-79. [A thoughtful, detailed lay description of an accident investigation with all its complexity and social ramifications.]
- Waddell G. A new clinical model for the treatment of low-back pain. *Spine*. 1987;12:632-644.

## **Occupational Injuries and Musculoskeletal Disorders**

- Waddell G. Low back pain: a twentieth century healthcare enigma. *Spine*. 1996;21(24):2820–2825.
- Webster BS, Snook SH. The cost of compensable low back pain. *J Occup Med*. 1990;32:13–15.
- Westgaard RH, Winkel J. Ergonomic intervention research for improved musculoskeletal health: a critical review. *Int J Ind Ergon*. 1997;20:463–500.

*This page intentionally left blank*

**THE PRAEGER  
HANDBOOK OF  
OCCUPATIONAL AND  
ENVIRONMENTAL  
MEDICINE**

*This page intentionally left blank*

# **THE PRAEGER HANDBOOK OF OCCUPATIONAL AND ENVIRONMENTAL MEDICINE**

---

**VOLUME**

**III** Practice Insights

Tee L. Guidotti, MD, MPH



AN IMPRINT OF ABC-CLIO, LLC  
Santa Barbara, California • Denver, Colorado • Oxford, England

## **Copyright Acknowledgments**

Material from two chapters adapted from Tee L. Guidotti. "Taking the Occupational History." *Ann Intern Med.* 1983; 199-4: 641-651.

Copyright © 2010 by Tee L. Guidotti

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, except for the inclusion of brief quotations in a review, without prior permission in writing from the publisher.

## **Library of Congress Cataloging-in-Publication Data**

Guidotti, Tee L.

The Praeger handbook of occupational and environmental medicine / Tee L. Guidotti.

p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-313-35999-6 (set hard copy : alk. paper)—ISBN 978-0-313-36001-5 (vol 1 hard copy : alk. paper)—ISBN 978-0-313-36003-9 (vol 2 hard copy : alk. paper)—ISBN 978-0-313-38204-8 (vol 3 hard copy : alk. paper)—ISBN 978-0-313-36000-8 (set ebook)—ISBN 978-0-313-36002-2 (vol 1 ebook)—ISBN 978-0-313-36004-6 (vol 2 ebook)—ISBN 978-0-313-38205-5 (vol 3 ebook)

1. Medicine, Industrial—Handbooks, manuals, etc. 2. Environmental health—Handbooks, manuals, etc. 3. Medicine, Industrial—Handbooks, manuals, etc. I. Title. II. Title: Praeger handbook of occupational and environmental medicine.

[DNLM: 1. Occupational Medicine. 2. Environmental Health. 3. Environmental Medicine. 4. Occupational Health. WA 400 G948p 2009]

RC963.32.G85 2010

616.9'803—dc22 2009046237

ISBN: 978-0-313-35999-6

EISBN: 978-0-313-36000-8

14 13 12 11 10 1 2 3 4 5

This book is also available on the World Wide Web as an eBook.

Visit [www.abc-clio.com](http://www.abc-clio.com) for details.

Praeger

An Imprint of ABC-CLIO, LLC

ABC-CLIO, LLC

130 Cremona Drive, P.O. Box 1911

Santa Barbara, California 93116-1911

This book is printed on acid-free paper 

Manufactured in the United States of America

# CONTENTS

## VOLUME 1: PRINCIPLES

<b>Preface</b>	<b>vii</b>
1. Occupational and Environmental Medicine	3
2. Toxicology	55
3. Epidemiology	201
4. Assessment of Exposure	295
5. Monitoring, Surveillance, and Screening	327
6. Safety and Injury Prevention	357
7. Risk Science and Risk Communication Skills	379
8. Prevention Science	459

## VOLUME 2: CENTRAL ISSUES

9. Physical Hazards and Conditions	505
10. Common Chemical Hazards	569
11. Buildings	689
12. Environmental Health	739
13. Stress and Psychosocial Factors at Work	809
14. Hazard Control	845
15. Clinical Evaluation	889
16. Occupational Injuries and Musculoskeletal Disorders	949

## VOLUME 3: PRACTICE INSIGHTS

17. Diseases and Health Protection	1013
18. Capacity for Work	1131

## **Contents**

19. Wellness and Health Promotion	1201
20. Occupational Health Services	1245
21. Workers' Compensation	1315
22. Emergency Management	1357
23. Medicolegal Services	1389
24. Regulatory Agencies	1443
25. Society, Critical Science, and Ethics	1481
26. Global Occupational and Environmental Health	1537
<b>Conclusion</b>	<b>1585</b>
<b>The Essential Library for the Occupational and Environmental Physician</b>	<b>1597</b>
<b>Index</b>	<b>1599</b>

# **VOLUME 3:**

## **Practice Insights**

*This page intentionally left blank*

# **17 DISEASES AND HEALTH PROTECTION**

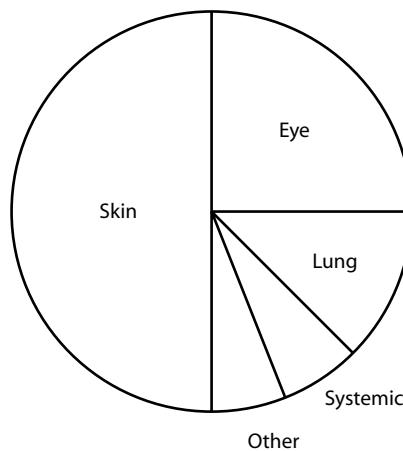
Diseases from occupational and environmental hazards represent failures of health protection. When a disease is discovered early, it may represent an opportunity for early treatment, but it always represents a warning that preventive measures are not working. The occupational and environmental medicine (OEM) physician, by virtue of medical knowledge, generally understands the disease process best and knows what to look for. The OEM physician is not necessarily an expert in treating these same diseases but does not have to be. That is what specialists are for.

Medicine is richly endowed with excellent clinicians, but there are few physicians who are effective in prevention and managing the social consequences of disease, which is the essence of the value that OEM brings to society. Oncologists concentrate on curing or managing cancer, not on assessing causation or preventing it, unless they have a personal interest and are motivated by commitment. OEM physicians concentrate on preventing and assessing the risk of causation of cancer and other diseases, not in treating them, unless they want to and are motivated by the rewards of patient care. Except for a few diseases that are uniquely occupational and require specialized skills to manage, such as lead toxicity in the adult

sufficient to require chelation, the value the OEM physician brings is in understanding the cause, promoting primary prevention, recognizing the disorder early, limiting disability, and mitigating the consequences, not providing unique or specialized treatment or procedures.

Occupational and environmental diseases range over the spectrum of clinical medicine, involving every organ system. A limited number are relatively common, are important because of their associations, and present critical issues for the OEM physician beyond patient care. These are emphasized in this chapter. This chapter also emphasizes occupational diseases over environmental because the workplace is the site of greatest risk.

Occupational diseases usually present more of a diagnostic dilemma than occupational injuries. In practice, the distribution of occupational diseases in North America can be approximated by the “rule of halves” (Figure 17.1), a general approximation that is useful in planning occupational health services. This rule of thumb states that the distribution of occupational diseases in a large working population in a diversified economy tends to be divided roughly as



**Figure 17.1.** The proportion of the major classes of occupational disease can be approximated by the “rule of halves,” which is useful for planning purposes.

follows: skin disorders account for roughly half of all occupational diseases, eye disorders roughly half of the remainder (or one-quarter), lung disorders half of that (or one-eighth), and half of the residual is systemic toxicity problems. The remainder tends to be a mix of many types of problems. This general approximation holds true as a rule of thumb for industrialized communities but may be distorted somewhat in smaller communities in which a single dominant industry presents an unusual hazard, such as coal mining.

Environmental diseases are generally more familiar to the physician outside OEM. At the level of regulatory control in most developed countries, overt disease caused by environmental exposure is rare. More common is aggravation of existing disorders, especially asthma, by environmental triggers, such as ozone or mold allergens. Environmental hazards are likely to play their most important role as risk factors in multifactorial diseases, such as the contribution of lead to behavioral disorders in children or of air pollution to low lung function. Environmental hazards are more likely in general to have an effect on children than adults.

In the discussion to follow, special emphasis is placed on occupational dermatoses, lung diseases, and toxic conditions because of their frequency, noise-induced hearing loss because of its frequency and the ease of prevention, and cancer because of its high frequency in certain occupational groups and its high visibility as an occupational health problem. Eye disorders are usually treated by specialists or physicians already trained for the practice and are not discussed in detail here.

Toxic disorders are an important class of occupational disease, but with a few specific exceptions, systemic toxicity is dealt with elsewhere in this book, usually where the agent is first introduced and discussed, generally Chapter 2 or 10.

This chapter does not provide detailed guidance on treatment. The reader is referred to the American College of Occupational and Environmental Medicine's *Occupational Medicine Practice Guidelines* (referred to as the *ACOEM Practice Guidelines*, or *APG*) or other authoritative references on management for treatment guidance.

## CLINICAL APPROACH

The practice of clinical occupational medicine seldom changes because of advances in medicine. Occupational medical practice changes little because of the introduction of new drugs and treatment. Rather, practice is changing with the changing economy, the introduction of new industrial processes, and changes in the workforce. The North American economy continues to move in the direction of domination by service occupations. The manufacturing sector frequently introduces new and different products and processes that create different workplace conditions.

Patterns of occupational disease change with the introduction of new and sometimes exotic chemicals and, even more importantly, the adaptation of chemicals that were frequently encountered in the past in new applications in which their hazards were not recognized. It is not uncommon in occupational medicine to see old hazards reappear in new workplaces. Arsenic, for example, was so closely associated with mining and smelting that it never occurred to technology wizards in microelectronics, with a very different mindset, that it would be a problem, as it became in the early days of the semiconductor industry. Indeed it is a truism of occupational medicine that exposures never truly disappear from the workplace—they reappear in new technologies, where they are often overlooked until they begin to cause problems.

Likewise, old hazards tends to reappear in the evaluation of new, or at least newly appreciated, risks as more is learned about mechanisms of disease. Previously, arsenic was known to cause groundwater contamination from natural sources and at those levels was associated with a risk of bladder, skin, and lung cancer. Now, arsenic is known to be associated with cardiovascular disease and neurocognitive impairment, as are mercury and lead.

Occupational medicine is more than the identification and treatment of diseases and injuries. It is fundamentally involved in the prevention of these disorders. Prevention, however, is usually easier and more readily undertaken if one has an understanding of what is

being prevented. This chapter will concentrate on the identification and understanding of important occupational diseases.

Five questions will be answered:

1. What are occupational disorders?
2. How does one recognize them?
3. How does one evaluate them?
4. What does one do when one finds them?
5. Who is available to help the physician in evaluating occupational disorders?

Occupational diseases are disorders that arise directly out of the workplace, working conditions, and exposures, including physical hazards that occur at work. By extension, they may include disorders that arise from work that is not part of an employment relationship, such as avocational work (hobbies or volunteer work), or from military or civil service, that share the same conditions and causes. The definition is important because there are other diseases that may be triggered, aggravated, or exacerbated but not caused at work; these diseases are usually called “work-related diseases.” Many other diseases may be important with respect to the capacity to work but are not caused by workplace conditions. These distinctions are very important with respect to diagnosis, recognition, compensation, and prevention. For example, asthma that arises *de novo* from sensitization to a chemical used on the job is an occupational disease, and failure to recognize this will put the patient at risk if exposure continues; it is compensable under workers’ compensation and can best be prevented by removing the antigen or reassigning the worker. Asthma that arises in a worker unrelated to their work may be made worse by exposure to irritants on the job, including exposures that would not bother most people; it is treated differently under many workers’ compensation systems, and in the individual case controlling the level of exposure may be all that is needed. Asthma that has no correlation with working conditions may be a health problem for the

worker and a serious cause of lost productivity in the workplace; it is not compensable under workers' compensation and at most would require an accommodation (in the United States, under the Americans with Disabilities Act).

Occupational diseases come to the attention of the physician through correlating the occupational history, which is discussed in considerable detail in Chapter 15, with clinical findings. The occupational history permits one to develop an exposure profile for the individual and in the format provided allows one to obtain information on the workplace, on hobbies, and on avocational exposure and exposure in the home.

Chapter 15 presents a guide for incorporating the occupational history and exposure profile into the clinical assessment and using resources available to physicians for interpreting the findings. This stage is comparable to the traditional medical process of deriving a differential diagnosis and then ruling in or out each possibility.

One evaluates occupational disorders by methodically considering each of five points:

1. Structural changes
2. Functional changes
3. Diagnosis and pathophysiology
4. Exposure circumstances
5. Behavioral causes (i.e., why exposure was allowed to occur)

Structural changes are often the first clue to the possibility of an occupational association with a disorder. For example, the evaluation of occupational lung disease often begins with changes in tissue structure as reflected in the chest film. Structural changes may be the only evidence of a lesion, and one may not be able to document functional impairment in the individual. This is usually the case, for example, with early stages of a pneumoconiosis.

Functional changes must be evaluated in order to assess the degree of impairment that may be present and to determine if the structural

change has progressed to the point of interfering with the worker-patient's ability to function, either on the job or in daily life. Examples of functional assessment are spirometry and exercise testing. At times structural changes are not at all evident or at least are invisible to our clinical methods of detection and in such circumstances functional changes may be the only way to document the presence of an occupational disorder. This is particularly true in the case of occupational asthma.

The medical diagnosis of an occupational disorder is often less important in occupational medicine than it might be in nonoccupational practice. Diagnosis is important to give an indication of prognosis and helps to define treatment. Other than systemic toxicity, however, most occupational diseases are treated generically, not with targeted therapy. Structural or functional changes may be more significant in identifying the proper category of occupational disorders, regardless of the specific diagnosis. On the other hand, in the occupational setting a diagnosis, however firm, is not sufficient in the evaluation of the patient. As a practical matter, causation analysis and functional assessment, which are usually subsidiary in general medicine, are at least as important as the clinical diagnosis. Causation is absolutely required because without a conclusion regarding etiology and the agent or exposure responsible, the link with work cannot be made, compensation cannot be considered, and prevention cannot be undertaken. The functional assessment is as or more important than a definite diagnosis because it establishes the basis for compensation and fitness to work. The diagnosis, however, does not necessarily have to be exact to guide treatment or to initiate a claim.

Identification of the specific agent or hazard that caused the disorder is only half of the description of etiology. Causation in occupational cases also requires a definition of the circumstances of exposure, with an indication of what the individual may have been exposed to and why that particular individual came into contact with the hazard. A comprehensive description of the exposure circumstances requires an understanding of the job requirements, the conditions of work, and at times, behavioral factors related to how the individual or co-workers

perform on the job and the policies of management. The level of detail that may be required parallels that of the fitness-to-work evaluation (See Chapter 18.) These issues are seldom described in as much detail in the evaluation of personal health problems.

Many resources are available to help physicians in interpreting the occupational history, evaluating occupational disorders, and establishing the diagnosis and functional assessment in treating these occupational disorders. In addition there are many other occupational health professionals, occupational health nurses, occupational hygienists, safety engineers, ergonomists, radiation health professionals, audiologists, toxicologists, epidemiologists, risk and liability control personnel, and vocational rehabilitation councilors who are usually only too glad to assist the physician in evaluating a case either on a consultation basis or more often in an informal basis by telephone. These occupational health professionals constitute a valuable resource for the physician. A little advance preparation in getting to know them and where they are in the community often becomes exceedingly useful in an emergency or when in sorting out a difficult case. Occupational medicine is one field of medicine in which the team approach works well and in which the physician cannot expect to do everything alone.

Environmental medicine has fewer resources to draw on than occupational medicine, because the field has not been as well defined. The member clinics of the Association of Occupational and Environmental Clinics ([www.aoec.org](http://www.aoec.org)) represent a resource for consultation in both occupational medicine and environmental medicine that is reliable and firmly in the mainstream of medicine. For children, the national network of Pediatric Environmental Health Specialty Units (described in Chapter 12) are essential sources of information and only a telephone call or email away.

## **DERMATOSES**

Occupational skin disorders are the most common type of occupational disease, comprising a third to half of all occupational diseases seen by physicians. Occupational skin disorders are much more

common than available statistics would indicate because they are unlikely to be brought to medical attention unless they become severe or uncomfortable. They are often seen in service industries, such as hairdressing, healthcare, and restaurants (among bartenders, dishwashers, and food handlers) as well as in manufacturing. While rarely life-threatening, occupational dermatoses may be quite distressing to the patient and are usually easily preventable. When severe they can be disfiguring, depressing, and a limitation on earnings.

The management of occupational dermatoses is similar to that of the same skin disorders in nonoccupational settings, with the additional feature that effective prevention and control depends on identifying the occupational cause of the disorder and protecting the worker from contact with the responsible agent. (See Chapter 14 for personal protection of the skin.)

The skin is a complex organ with a massive surface area and is uniquely accessible to examination. Although the skin provides an effective barrier to most physical and chemical exposures, it is not completely impermeable and represents a major pathway of absorption. (See Chapter 2.) The skin presents a number of host defenses that resist local injury, including buffering of pH within a narrow range, immune mechanisms, cooling by vasodilation, mechanical protection against abrasion (by the stratum corneum), protection against ultraviolet radiation (by the pigmented melanocytes of the basal cell layer), and antimicrobial activity (fatty acids that inhibit bacterial and fungal growth on the skin surface).

The diagnosis of occupational dermatoses begins with the history and inspection. The changes are generally easily visible and often appear sharply delineated in areas of the body uncovered by work clothes: neckline, wrist, or face. The history of the reaction is important, as local contact dermatitis may generalize in the “id” reaction, involving distant sites not obviously related to the circumstances of exposure. Occasionally, usually in oil acne, the eruption may be on parts of the body covered by saturated clothes or occluded by heavy clothing in an environment of heat and humidity. The inside of work gloves may become contaminated during careless use, producing a

dermatitis on the hands despite a history of using personal protection. The history of the initial response of the rash may be useful if it is clearly associated with job duties or time in the workplace; often, however, recurrent dermatoses become relatively persistent and may not go away with time off work. It may take weeks for a severe contact dermatitis to heal after exposure ceases. The medications used to treat the dermatitis occasionally may aggravate it or change its presentation.

Factors that aggravate occupational dermatoses, or make it more likely that an exposure will result in a dermatitis, include heat, moisture, physical abrasion, pressure, and occlusion. All these factors may be present in the case of tight gloves or ill-fitting work clothes. Excessive dryness may also cause skin problems. Some measures ostensibly taken to protect the skin may make the dermatitis worse. Frequent hand washing, for example, may cause chapping, cracking, loss of oils, and exacerbated dermatitis.

Prevention of occupational dermatitis rests on good hygiene and the use of personal protection, including impermeable but cool and comfortable gloves, kept clean on the inside, and work clothing, including aprons. Barrier creams are oil-based substances applied to exposed parts of the body to prevent irritation of the skin. They are marginally useful but quickly lose their protective action with time and activity. Barrier creams can be used within gloves for added protection, especially if the glove itself is thought to be part of the problem. (Personal protection is described in Chapter 14.)

Treatment of occupational dermatoses is specific to the type of dermatitis and similar to that of other dermatitides. Control of exposure to irritants and allergens and any aggravating factors is the key to management, however. Treatment is seldom very effective if exposure and irritation persist.

## **Acute Skin Injury**

Acute irritant skin injury resembles a burn and can be exceptionally severe (see Chapter 16). These “chemical burns” occur as a consequence of contact with strong acids (for example, plating

solutions), strong bases (for example, cement or ammonia), or highly reactive chemicals (for example, ethylene oxide or potassium permanganate). Prolonged immersion in solvents, which extract lipids from the skin, may cause a similar clinical picture, often accompanied by cracking and fissuring. The chemical burn usually begins with erythema, pain, and pruritus and may develop blisters or ulcers. Injury in these cases penetrates to the dermis and may cause severe scarring. Recovery may be accompanied by disfiguring hyperpigmentation in persons with pale skin or either hyper- or hypopigmentation in persons of color. Repeated exposure to more dilute strong irritants or to solvents may result in chronic skin changes, either resulting in a hypertrophic dermatitis or skin atrophy and nummular dermatosis, which is characterized by skin hypertrophy.

One strong irritant deserves special mention because of its extreme hazard. Hydrofluoric acid (HF) is heavily used in the semiconductor industry for etching silicone chips and removing surface oxides, in the oil industry as a fracturing agent and for maintenance of reaction vessels, in metallurgy to remove impurities and oxides from steel (a process known as “pickling”), and as a fluorine source in making fluoropolymers. HF is a weak acid in chemical terms, but when the acid dissociates it forms fluoride ion ( $F^-$ ), which has a strong affinity for calcium and penetrates cell membranes very efficiently. An HF burn causes deep and spreading tissue injury that is very difficult to contain and manage. HF burns begin with erythema and swelling and, while they may not initially look very bad, may progress over hours to blanching (extreme pallor), blistering, and excruciating pain, out of all proportion to the visible sign of injury. HF burns may easily lead to necrosis and extensive tissue loss, and because of fluoride’s affinity for calcium in bone they may demineralize bone underneath the skin destruction, which can result in loss of fingers and toes. Amputation is sometimes required. Treatment, to be initiated as early as possible, is calcium gluconate as a surface gel (which has to be prepared at the time using KY Jelly ® or a similar product because it is no longer commercially available), as a dilute solution for eye irrigation, by tissue infiltration within and beyond

the obviously affected skin area, and if necessary by intra-arterial injection, which may save body parts. Any OEM physician who could possibly be called upon to manage an HF burn is advised to receive special training in their treatment (provided at no cost by manufacturers of HF) and to maintain an appropriately large supply of calcium gluconate at the ready.

## CONTACT DERMATITIS

“Contact dermatitis” comprises 75 percent of occupational skin disorders. Of this, approximately 80 percent is due to local irritation and 20 percent is due to an allergic response to specific antigens. Atopic individuals are much more likely (by 13–14 times) to experience contact dermatitis of either type, not just allergic contact dermatitis, than nonatopic individuals. Both allergic and irritant contact dermatides are aggravated by heat, humidity, and physical abrasion.

An “irritant contact dermatitis” may take many forms, but usually presents as eczema or changes in pigmentation. The irritation may come from relatively prolonged contact with a weak irritant, which may be almost any compound in the workplace. The most common irritants are acids, bases, solvents, soaps, and detergents. The latter four tend to extract oils and lipid from the skin, creating a dry, chapped rash. Once the surface of the skin is broken, a rash tends to perpetuate itself, especially with frequent hand washing. Agents used to cleanse the skin may perpetuate the irritation. Abrasive (usually pumice-containing) soaps, waterless hand cleansers, and solvents are particularly troublesome in this regard and should be avoided once the dermatitis appears. An irritant contact dermatitis frequently results from the cumulative effect of several irritating exposures. Although sometimes refractory, irritant contact dermatitis can usually be managed by reducing the sum total of irritant exposures and often allows continuation of work activities with personal protection or reduced exposure.

“Allergic contact dermatitis,” or eczema, is a Type III immunologic reaction to a specific antigen. Common sensitizers include nickel, disulfiram (a rubber constituent), epoxy and other resins,

and organic dyes. Once an individual is sensitized to an antigen, minimal contact thereafter produces a reaction. This makes management much more difficult, in general, than for irritant contact dermatitis and almost always requires avoiding the antigen entirely. Fortunately, antigens encountered in workplaces tend to be relatively weak and often require considerable time to provoke sensitization. A dermatitis occurring within the first month on a new job is unlikely to be allergic in nature. This is important because effective control of allergic contact dermatitis is often more difficult than that of an irritant effect if the worker cannot avoid contact with the sensitizing agent in the workplace. A rash occurring several weeks to months after beginning a new job or after the introduction of a new chemical into the workplace may be either an allergic or irritant mechanism. Sensitization to perfumes and cosmetic constituents sometimes produce an allergic dermatitis that may appear at home or at work, confusing evaluation. Persons with a history of eczema may be at risk for the id reaction, in which the dermatitis generalizes to involve previously normal skin, often in areas distant from the initial eruption. This can create a very confusing clinical picture.

“Patch testing” to evaluate allergic contact dermatitis is a useful tool in occupational medicine with appropriate standardization of practice and in experienced hands. Patch testing is an important but limited diagnostic tool in the evaluation of occupational dermatoses. The kits, or “trays,” with which patch testing is conducted do not contain all possible occupational antigens for every workplace situation and are more likely to miss an antigen than to register a false positive. Thus, patch testing should normally be considered confirmatory if positive rather than definitive evidence against an occupational association if negative. It is useful in identifying specific allergens to which the worker has become sensitive. In theory, exposure to that substance can be avoided subsequently. The antigen is supplied in a diluted solution that is applied to the skin in the form of a saturated piece of gauze or cellular discs, under an occlusive dressing, usually on the worker’s back in groups of 23. Localized

eczematous reactions occur in the presence of allergy within 48–72 hours in most cases or up to 96 hours in a minority of reactions.

Patch testing is conducted following a set of strict international guidelines using standard solutions standardized by the American Academy of Dermatology. It is possible to formulate solutions of allergens not included in the standard kits, especially when the material is water-soluble or can be obtained in the form of metal salts. There are serious limitations to patch testing as a diagnostic procedure, however. False positive tests can occur on highly reactive skin and with the application of substances that are also irritants. Workers can become newly sensitized to allergens during the course of patch testing to chemicals they tolerated previously. False negatives can occur, especially if the worker is under treatment or the solution is not standard. Patch testing only indicates the presence of a specific allergic reaction; it cannot rule out an irritant contact dermatitis or, indeed, an allergic contact dermatitis resulting from exposure to an allergen that does not happen to be included in the panel used. Patch testing should be performed in a suitably equipped facility by a dermatologist or allergist familiar with internationally standardized protocols and with the limitations of the technique. Because it is time-consuming and often inconclusive, patch testing is not performed by many dermatologists and may be hard to arrange.

Figure 17.2 presents a characteristic example of allergic contact dermatitis in a thirty-two-year-old man at a pressed wood plant. He developed this eczematous rash from exposure to resin dust compounded with glue containing glutaraldehyde; subsequently he also developed occupational asthma with wheezing following exposure. Patch testing showed a 3+ reaction to glutaraldehyde.

“Contact urticaria,” resulting from sensitization to allergies involved by skin absorption, is a relatively uncommon reaction expressed as local hives. Generalized urticaria can occur after inhalation of an allergen, usually in the context of a respiratory reaction. However, urticaria often occurs with no cause identified.

Strong irritants, such as acids, alkalis, and highly reactive chemicals such as ethylene oxide, are not as likely to cause contact dermatitis as they are to cause an obvious chemical burn, blistering, or ulceration.



**Figure 17.2.** Allergic contact dermatitis in a thirty-two-year-old worker sensitive to glutaraldehyde in the glue used in a pressed wood plant. He also had occupational asthma, with bronchospasm following exposure. Note the distribution of the rash in the exposed area between glove and sleeve.



**Figure 17.3.** Depigmentation and hypertrophic skin changes in the hands of a fifty-three-year-old spray painter who had frequent skin contact with solvents.

Chronic or persistently recurrent dermatitis may lead to secondary effects, such as hypo- or hyperpigmentation and hypertrophy. Figure 17.3 shows the effect of chronic, recurrent solvent exposure on the hands of a spray painter.

## Other Dermatoses

Other dermatoses are noteworthy as clues to an occupational exposure. Photosensitization may appear as an acute sunburn or eczema on exposed areas of skin, often accompanied by conjunctivitis. This occurs most commonly in occupations involving heavy exposure to coal tar products, such as roofing, but has also been seen in workers exposed to light-cured acrylic and vaporized epoxy resins and among bartenders, pickers, and grocery store workers who handle produce, principally celery and limes, that are bruised and infected with a common fungus which secretes a type of psoralen that is photosensitizing.

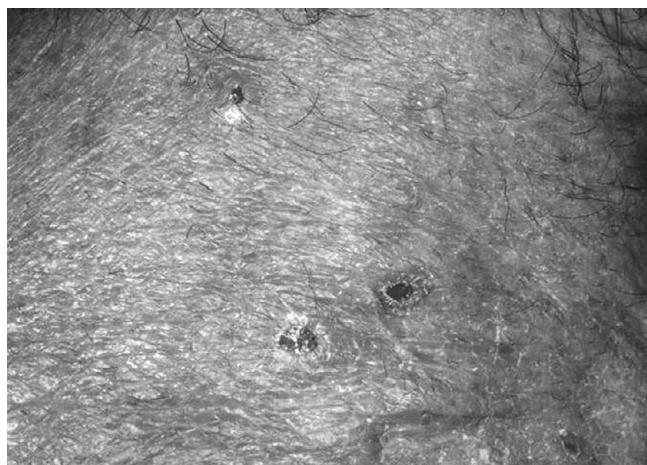
“Oil acne” or “oil folliculitis” is often associated with exposure to greases and lubricating oils in settings in which washing is inconvenient or work clothes become saturated. It usually appears on the hands and forearms because of exposure in these areas. Hair follicles are often inflamed and occluded as a first sign.

“Chloracne” is a persistent and often severe form of nonpostular acne, usually affecting the face, especially in the preauricular area. Chloracne reflects systemic exposure to chlorinated cyclic hydrocarbons (e.g., pentachlorophenol, PCBs, furans, and dioxins). Exposure to these agents usually occurs in one of a few specific situations: an industrial accident involving chemical reactions, poor manufacturing processes, application of wood preservatives, burning of preserved wood, and exposure to chemical wastes. Unlike oil acne, chloracne is a sign of significant systemic toxic exposure and may be associated with other effects of exposure, such as hepatic disorders and neuropathies. It tends to be refractory to treatment. Suspicion of chloracne merits referral to a dermatologist, evaluation by a toxicologist or OEM physician, and close follow-up.

A common ulcerating occupational dermatitis is a reaction to chromium. Typically, “chrome ulcers” present as punctate, persistent open sores on the hands that are very slow to heal. They may appear in workers handling cement, which contains small amounts of chromium (see Figure 17.4).

Skin cancer is a serious occupational health problem, mostly related to work out of doors. The nature and distribution of skin cancers among workers at risk parallels that of the general population and reflects primarily exposure to ultraviolet light. The risk of skin cancer, including melanoma, can be reduced significantly by protecting clothing and the provision of shelter for workers outdoors. Photosensitizers, such as coal tar products (which are concentrated polycyclic aromatic hydrocarbons), play an important role in some industries, enhancing the initiating effect of UV radiation.

Systemic exposure to arsenic is associated with basal cell carcinoma and squamous cell carcinoma, both invasive and *in situ* (known as Bowen's disease). These may be associated with the characteristic nonmalignant skin rash of arsenic, hyperkeratotic skin changes that are usually most obvious on the palms and plantar surface.



**Figure 17.4.** Chrome ulcers on the hands of a man who handled dirt contaminated with potassium chromate; they are usually associated with handling cement.

Numerous specific skin disorders are associated with particular occupations; others are cutaneous signs of systemic disease that may have an occupational cause. Agricultural, food, and forestry workers are particularly susceptible groups for dermatoses. Causes of dermatitis in these occupations may include sensitization to certain pesticides, photosensitization (as noted above), irritation by contact with poison oak or ivy, insect bites, or exposure to common solvents and chemicals. Infectious occupational dermatoses may also be seen in forestry, agriculture, food, and forestry workers, including Lyme disease, sporotrichosis (classically seen in horticulture after the skin is pricked by rose thorns), and erysipeloid. The clinical presentation of dermatoses in agricultural and forestry workers is often modified by sun exposure, heat and humidity, and the improper use of protective equipment, which may actually retain sensitizing or irritating agents next to the skin if the inside of gloves, masks, or clothing becomes contaminated.

Occupational dermatoses may also be caused by mechanical factors, such as friction (calluses, abrasions, exacerbation of psoriasis), pressure (blistering, ulcers), and minor trauma. Exposure to vibration, heat, and cold may result in specific cutaneous manifestations, as described in Chapter 9. Penetration by foreign bodies, including fibrous glass, asbestos fibers, or metal fragments, often leads to granuloma formation and palpable nodules. Recovery from inflammatory conditions often leads to pigmentation changes and may be cosmetically troublesome.

## **RESPIRATORY DISEASES**

Table 17.1 summarizes the common benign and malignant occupational and environmental diseases of the respiratory tract and thorax most likely to be encountered in OEM practice. As a practical matter, occupational lung disorders are most critical to understand, and the rest of this section will be devoted primarily to lung disorders and primarily to occupational hazards because the greatest hazard is in the workplace. However, it should be recognized that

upper airway disorders occur commonly, especially associated with atopy and irritation, and may be associated with the usual symptoms of cough, rhinitis, and sometimes voice changes.

Diseases of the pleura, because of its inaccessibility except by lymphatic transport, occur primarily in response to asbestos.

Within the lung, the disorders of greatest concern in OEM are the “pneumoconioses” (dust-related diseases of the lung), airway disorders (asthma or bronchitis and their variants), “hypersensitivity pneumonitides” (a class of immune-mediated responses to inhaled antigens), “toxic inhalation” (an acute syndrome associated with

**Table 17.1.** Disorders of the Respiratory Tract in Occupational and Environmental Medicine

	Benign	Malignant
Upper airway (except larynx)	Rhinitis, sinusitis, septal perforation	Nasal and sinus cancer, pharyngeal cancer
Larynx	Laryngitis, dysphonia, stridor, obstruction, polyps	Laryngeal carcinoma
Lower airways (bronchi, bronchioles)	Bronchoconstriction, asthma, bronchitis, chronic obstructive airway disease, bronchiolitis	Bronchogenic carcinoma
Parenchyma, (alveoli)	Pneumoconioses, emphysema, hypersensitivity pneumonitis, pneumonia	(Usually metastatic)
Thorax (pleura)	Pleural fibrosis (plaques and diffuse thickening), acute pleural effusion of asbestos, “rounded atelectasis” syndrome	Mesothelioma, metastatic disease

diffuse alveolar injury), and cancer of the lung and pleura. “Inhalational fevers” (acute, self-limited febrile illnesses associated with inhalation of zinc or copper) are characterized by a transient infiltrate but are really more transient systemic toxicities than lung disorders. (Inhalation fevers are discussed briefly in Chapter 10.)

### Expressions of Lung Injury

The lung has a limited repertoire of change. The patterns of expression of lung injury are limited by the structure and functional possibilities of the organ. Occupational and environmental lung diseases mimic many if not most nonoccupational disorders. Depending on the mechanism of the response to the inhaled toxic agent, the disease may result in a restrictive or an obstructive defect and a malignant or a benign process and will primarily affect the airway or the parenchyma. Some occupational lung disorders have extrapulmonary manifestations even in the absence of malignancy. Indeed, the range of structural and functional abnormalities produced by occupational exposures embraces most of the clinical spectrum of pulmonary medicine.

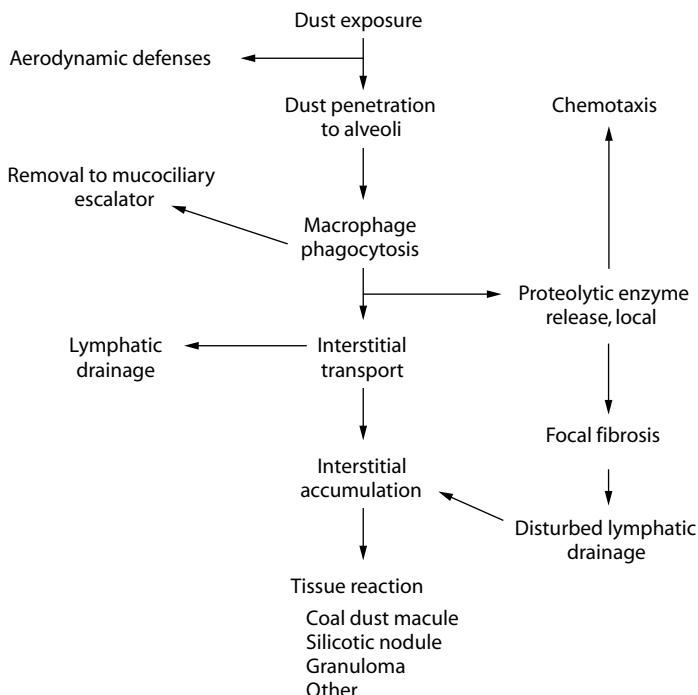
The lungs are highly vulnerable to injury from the dusts and gases in the atmosphere. To protect the delicate structures of the lower respiratory tract, the human lung is protected by elaborate defense mechanisms. These defense mechanisms may be compromised or overwhelmed by occupational exposures, leading to various types of occupational lung diseases.

Toxic agents that may be inhaled can be categorized as either particulate (solid) or nonparticulate (gas). Toxic exposures in the workplace are often a mixture of gaseous and particulate compounds, as in exposure to welding fumes or combustion exhausts. In many cases gaseous compounds are carried adsorbed onto the surface of particles. In such dual exposures the toxic substances on the particle are carried much more deeply into the lower respiratory tract than they would otherwise penetrate. The clinical expression of the toxic reaction is determined by the dose

received, the pathophysiology of the response, characteristics of the host, and the presence of modifying factors such as other exposures and infectious agents. Chapter 2 discusses inhalation toxicology in more detail.

Inhaled dust is rapidly cleared by the highly effective centripetal mucociliary stream, which removes a prodigious burden. A relatively small quantity of residual dust is left behind for phagocytosis by alveolar macrophages. The macrophages, having engulfed the individual particles, may migrate on the alveolar surface or interstitially to the edge of the mucociliary escalator (this process is not well understood), or they may migrate interstitially. Although interstitial migration probably drains into lymphatic system for the most part, a significant amount of dust is trapped in local concentrations at the coal dust macules, where presumably drainage is impaired or overwhelmed or reaches an anatomic "dead end." These transport mechanisms may be impaired by inflammatory respiratory disease, by simultaneous exposure to irritant exogenous agents, which may be ciliostatic, or by exposure to other particulates, which may saturate the phagocytic capacity of the macrophages.

Macrophages release proteases and inflammatory products that cause local damage and initiate a dysfunctional repair mechanism by fibroblasts. Phagocytosis of even inert particles, such as latex beads, by alveolar macrophages is enough to release small quantities of lysosomal enzymes into the surrounding milieu. Dusts that are refractory to digestion and that initiate cytotoxicity, such as silica, cause more release. These include collagenase, elastase, plasminogen activator, and lysosomal hydroxylases. Thus, alveolar macrophages may not be activated by ingesting inert particles such as carbon, but they are stimulated to release, inadvertently as it were, potentially locally destructive enzymes. This is reasonable in a system evolved to dispatch microorganisms, but it is counterproductive in dealing with inert particles. Furthermore, partially denatured protein is chemotactic, so that the cellular reaction perpetuates itself. The schema in Figure 17.5 illustrates the



**Figure 17.5.** Fate of inhaled particle penetrating to lower respiratory tract.

common mechanism for reacting to dust deposition in the lung, in this case the formation and extension of the coal macule in coal workers' pneumoconioses.

Focal fibrosis and focal emphysema may result, which contributes to an obstructive component of dust-related disease. There are opportunities for individual variation implied by this schema. Miners with  $\alpha_1$ -antitrypsin deficiency, for example, will have more extensive local destruction than otherwise, but this combination is obviously rare. The avidity and efficiency of macrophages in clearing mycobacteria have a genetic component.

Nonparticulate, gaseous toxic agents penetrate into the respiratory tract depending on how soluble they are in water. The mucosa of the respiratory tract absorbs and removes most readily toxic gases that are water soluble. Host defense mechanisms against infection

and particulates, especially the mucociliary escalator and alveolar macrophage, are very vulnerable to the effects of toxic gases, and their efficiency is often considerably reduced after exposure. Cigarette smoking exposes the lungs to the toxic agents in cigarette smoke, interferes with the mechanisms of host defense, and often interacts with other toxic exposures to produce more severe toxicity than could otherwise occur.

### **Clinical Assessment**

The evaluation of occupational lung disease must address five essential questions:

1. What is the nature of the process?
2. What exposure in the patient's employment may have been responsible?
3. What is the prognosis of this condition and what ultimate level of disability accompanies the natural history of the disorder?
4. What can be done to control or limit the disease process?
5. Are other persons in the same workplace likely to be affected, now or in the future?

The first question, on the nature of the lesion, requires a description of the structure, function, and malignant potential of the process. Structure is usually evaluated by chest film (increasingly supported by high-resolution CT scanning), function by pulmonary function tests, and malignant potential by biopsy or cytopathology. The radiological findings provide evidence of the nature of the process, such as fibrosis, when it affects the lung's gross structure. Few occupational lung diseases have characteristic findings on chest film, although asbestosis and silicosis do have pathognomonic signs, as will be described. Functional impairment, such as airway obstruction, can be assessed by history or best by spirometry. Determination of malignancy requires tissue or cytologic confirmation, although radiological evidence may be very convincing.

The second question, concerning exposure, is often apparent from the occupational history. A history of exposure to a known carcinogen, a fibrogenic dust, or a potent sensitizing agent should be considered for significance in light of the presenting symptoms and the known effects of the agent. Identification of the agents helps to suggest a prognosis and is essential for establishing eligibility for compensation. Defining the cause may contribute to future preventive measures. The clinical management of individual patients may or may not be affected by identification of the agent.

The third question, on prognosis and disability, depends on the agent responsible and on the functional status of the patient. Silicosis and asbestosis, the two most common pneumoconioses of the over 200 known, both produce serious restrictive disease that will often advance regardless of future exposure. Most other pneumoconioses produce structural abnormalities on chest film greatly out of proportion to the modest changes they cause in pulmonary function and usually stabilize or may even regress after exposure ceases, presumably because of mobilization of dust and remodeling of fibrosis.

The fifth question, on the risks to other workers, is critical. Other employees who worked in the same plant may be at risk for the same outcome. The physician should always inquire about the health of co-workers, the number of employees in the workplace, and whether the hazard is still present.

## **Pneumoconioses**

The pneumoconioses are diseases characterized by the deposition of dust in the lung and the pulmonary response to its presence. As is the case with tuberculosis, the lung's intrinsic reaction is a key part of the development of the disease. Usually, this consists of an inflammatory reaction, usually low-grade, that may or may not lead to extensive fibrosis. The degree of fibrosis that results varies with the properties of the dust. Silica and the fibrous silicates, such as asbestos or zeolite, cause intense fibrotic reactions. Carbon black or iron oxide provoke only small and localized reactions. Coal dust, the other

common pneumoconiosis, falls in between depending on the composition of the coal. Some pneumoconioses are also associated with immune responses that modify their presentation and add new mechanisms of action: silica (which has a strong association with autoimmune disorders), beryllium (which produce a sarcoidosis-like response), and hard-metal disease (which, as a result of the cobalt content of the alloy, is associated with occupational asthma). Others may be associated with direct toxicity if the dust is soluble.

Pneumoconioses caused by metals of high atomic number (which are dense to x-rays) present an exaggerated appearance on chest film. They usually appear first as dense, multiple-nodular opacities. Correlation between radiological patterns and functional impairment is often poor. Function is usually preserved, as measured by routine spirometry, until the disease becomes advanced. The usual pattern of functional impairment in an advanced pneumoconiosis is restrictive, but mild obstruction may occur, especially when there is also a bronchitis. The identification of a pneumoconiosis and its differentiation from idiopathic interstitial fibrosis, sarcoidosis, the interstitial pneumonias, or cancer is sometimes possible only by open biopsy. It is important in such cases to notify the pathologist receiving the tissue when a pneumoconiosis is expected. Special techniques are often required to identify the type of dust in a pneumoconiosis and the diagnosis can be overlooked even by experienced pathologists.

The evaluation of a suspected pneumoconiosis begins with the chest film. The simple PA chest film is a clinical tool of remarkable usefulness, particularly in the evaluation of occupational lung disorders. Most pneumoconioses appear first as multiple, rounded nodules in the lower lung fields. Silicosis, although it presents first as rounded nodules, occurs first in the upper lung fields. Asbestosis presents irregular nodules. Otherwise, the pneumoconioses have a relatively uniform appearance that resembles sarcoidosis.

Over the years, experience with the radiographic diagnosis of occupational lung disease has led to the appreciation of perceptual problems on the part of film readers and their resolution by

adopting certain conventions. The standard system for communicating the morphology, nodularity, and severity of radiographic changes in cases of suspected pneumoconiosis is the so-called "ILO Classification of the Radiographic Appearance of the Pneumoconioses." (ILO stands for International Labour Office.) This system is based on a psychological process of dealing with uncertainty in the mind of the physician reading the chest film. This necessarily involved system uses four basic categories expanded into a twelve-point scale reflecting and actually using to advantage the uncertainty of classification in the mind of the reader. For example, for a film that is probably category 2 but for which the reader considered assigning category 1, the reader would indicate 2/1. The biggest difficulty with the ILO system comes in distinguishing between 1/0 and 0/1, since this represents the dividing point between presumed normal and presumed diseased. The system also provides an interpretation by the type of density observed, assigning a series of small letter designations depending on the shape, regularity of margins, and size of the densities. The system was designed and is used for classifying films for epidemiological purposes but has been adopted in the United States for certain purposes in determining eligibility for compensation. It is not appropriate for use in clinical diagnosis, since its use requires that the diagnosis already be made, but it is useful for staging. The ILO Classification, and most definitely the "B-reader program," which is the testing program managed by the National Institute of Occupational Safety and Health (NIOSH) for proficiency in applying it, is now based on obsolete technology. The conventional PA chest film on which it is based is rapidly becoming obsolete and even unavailable in many centers, replaced by digitized films. Guidelines for application of the ILO Classification to digital films are now available and a complementary classification system for HRCT (high-resolution computed tomography) has been proposed but is not universally accepted.

Pulmonary function in the pneumoconioses, generally, is mildly irritating for most dusts and may contribute to obstructive lung disease. For the fibrogenic dusts that are the major cause of serious

disease, function varies with the stage of the disease. In the earliest stages of most pneumoconioses, there may be mild obstructive changes due to airway irritation, even for benign dusts such as graphite. As the disorder progresses, some component of airway obstruction progresses due to focal emphysema, but it is almost always masked and overwhelmed by a progressive restrictive change due to the advancing fibrosis.

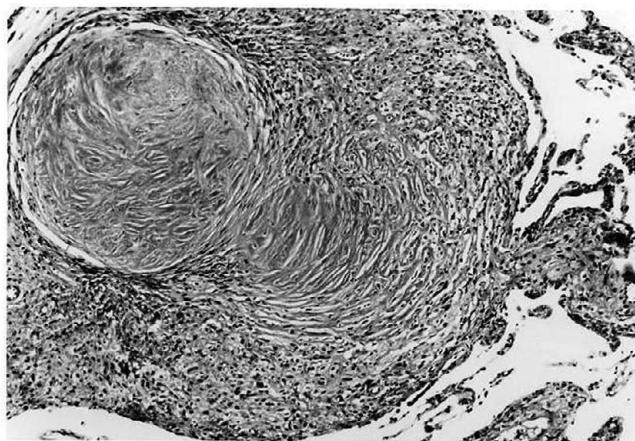
### ***Reactions to Fibrogenic Dusts***

The most important pneumoconioses in terms of frequency, medical management, and poor prognosis are the fibrogenic dusts, which induce scarring in the lung; the pneumoconioses arising from exposure to asbestos (“asbestosis”), crystalline silica (“silicosis”), and coal dust (“coal workers’ pneumoconiosis”) are historically and still today by far the most common fibrogenic dusts to which workers are exposed. The dusts responsible are minerals sharing silica as a common chemical element but with very different composition, chemical and physical properties, and potency in causing fibrosis in lung tissue. Silica is silicone dioxide in various forms, usually nearly pure  $\alpha$ -quartz in crystalline form. Noncrystalline silica is not so potently fibrogenic. (Hereafter in this section, “silica” will refer to crystalline silica.) Asbestos, which exists in many forms, is a silicate, with additional constituents. Coal dust is a complex mixture of predominantly carbon with some mineral composition and is more potently fibrogenic the greater the proportion of silica, although its pathology is distinctly different.

In each case, the dust, once inhaled, is picked up by defensive cells of the body, specifically the alveolar macrophage. The alveolar macrophage has as its mission to search out and destroy foreign and potentially threatening bits of matter, such as bacteria. It does so by a variety of physical and chemical reactions more appropriate to killing bacteria than digesting particles of mineral origin. The incidental release of protease and intracellular oxidizing agents damages the surrounding lung tissue.

The effects of the inhaled dust on the macrophage are also a critical aspect of its toxicity. In the case of these two dusts, the reactions of the macrophage to the dust create more damage than the dusts themselves might otherwise. The macrophage is itself killed or severely damaged by its encounter with the silica dust particle, by mechanisms that are still under investigation. The damaged cell, directly or indirectly, stimulates fibroblasts, the cells that lay down collagen, the rope-like substance that is the basic element of connective tissue, in an exuberant and abnormal way, resulting in scarring and stiff abnormal tissue in the interstitium, the structural tissue of the lung, between the walls of the alveoli. However, the pathologic features of the three dust diseases are very different.

In silicosis, the dust and its accompanying macrophage find their way directly and by lymphatic channels into the interstitial space, which tends to run alongside blood vessels in the lung. The scarring that takes place occurs in the interstitium in the form of a mass of rope-like fibers that thicken and stiffen the formerly delicate interstitial tissue. In addition, collections of dust-laden macrophages and certain lymphatic drainage areas become centers of particularly active



**Figure 17.6.** An exceptionally well-formed silicotic nodule. (Photograph courtesy of Dr. Jerrold L. Abraham.)

fibrosis. This is the origin of the “silicotic nodule,” a characteristic feature of silicosis not found in asbestosis or any other pneumoconiosis (Figure 17.6). These central areas thicken, coalesce with adjacent nodules, and become the centers of an extensive web of fibrosis that radiates out from the nodules. As the nodules grow larger, the tissue at the centre dies from ischemia, or lack of blood-born nourishment. This results in a thickened, almost glassy appearing center in the middle of the nodule, which otherwise resembles a fingerprint due to the whorled tangle of rope-like collagen fibers.

There is a characteristic lesion of coal workers’ pneumoconiosis, the coal dust macule, but the fibrosis is not as exuberant and the fibrosis is not organized in a concentric pattern as in silicosis.

By comparison, asbestosis is a ragged and irregular disease, resulting in extensive patches of fibrosis in the interstitium. The macrophage bearing an asbestos particle is greatly impeded in its ability to pass through the interstitial space and lymphatic channels compared to the macrophage bearing silica particles. That this would be so can be understood by simple geometry. Asbestos particles are long and thin and may get stuck while passing through narrow channels. Silica and coal dust particles, although irregular, are more nearly spherical. The contrast is like a person getting onto a crowded escalator carrying a beach ball (a particle of silica or coal dust) or carrying a ladder (asbestos). Many more asbestos particles get stuck in the interstitium. When they do travel considerable distances in the lymphatic channels, they cause considerable fibrosis wherever they tend to collect. The pleural surface, the tissue lining the thoracic cavity and its viscera, tends to trap asbestos fibers and so is particularly subject to this injury, leading to fibrosis in the form of pleural plaques and the risk of cancer.

### ***Silicosis and Silica-Related Disease***

Silicosis is an ancient disease that continues to occur today in numerous occupations, although it is entirely preventable. Worldwide, silicosis is the most prevalent serious occupational lung disease,

mostly in less economically developed countries where the resulting illness is a cause of significant disability and an economic drag on development. Although the incidence of obvious silicosis has declined in developed countries and it has been effectively eliminated in Western Europe and Canada, there are still over one hundred deaths every year from the disease in the United States and much unrecognized morbidity because of underreporting and lack of recognition of silica-related diseases beyond classical silicosis.

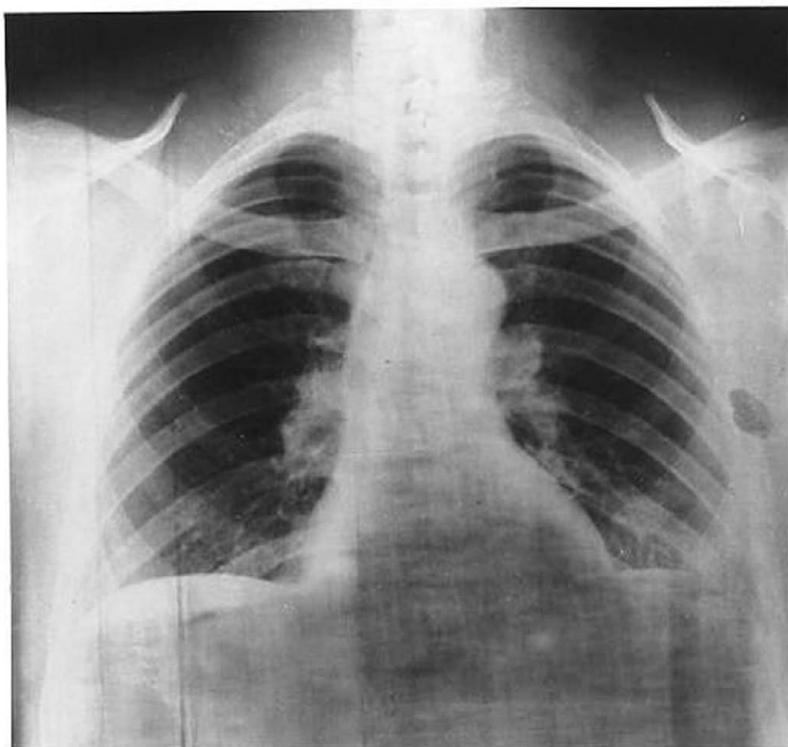
Silica dust is intensely fibrotic, producing a characteristic microscopic lesion called the “silicotic nodule,” a nidus of focal fibrosis surrounding a lymphatic deposition of silica dust, each of which may grow into a large, whorled fibrotic mass and coalesce with one another. The chest film appearance begins as isolated opacities against a normal lung parenchyma that progresses over years to a reticulonodular infiltrate and further progresses over months and years to severe fibrosis, formation of bullae, fibrotic masses, and calcification of the lymph nodes and parenchymal masses (see Figure 17.7). Two pathognomonic radiological features of severe silicosis are calcification of enlarged hilar nodes in a thin oval or elliptical pattern (“eggshell” calcification) and retraction and fibrosis of the parenchyma at the apices into dense crescents concave toward the hilum (“angel wings,” a mordant reference not only to the sign’s appearance but to its significance as a harbinger of death). Silicosis can be characterized by radiological stages and by variants characterized mostly by the rate of progression.

“Simple silicosis” is characterized by isolated rounded opacities, usually first visible in the upper lung fields. As it progresses, simple silicosis is characterized by proliferation of rounded small opacities resembling miliary tuberculosis or sarcoidosis (with which it is frequently confused), gradually increasing in density and dispersion among the lung fields (Figure 17.7). The effect is something like white sand scattered on the chest film. The parenchymal lesions may be associated with hypertrophy of the hilar lymph nodes, which later calcify around the periphery. At times, the lymph node hypertrophy can be extreme, and rare cases may not show parenchymal opacities on the chest film at all early on. (Figure 17.8 demonstrates such a



**Figure 17.7.** Early simple silicosis. (Reproduced with permission of the American Academy of Family Practice.)

case, which resembles sarcoidosis; the diagnosis must be made by biopsy.) There is often a thin dense ring of calcification around the lymph nodes—so-called egg shell calcification characteristic of silicosis that is only rarely seen in other conditions such as sarcoidosis. Progression of the disease is associated with increases in



**Figure 17.8.** Silicosis, biopsy proven, presenting as hilar lymphadenopathy.

the size and number of opacities (see Figure 17.9). Simple silicosis is usually not associated with significant respiratory impairment.

In “complicated silicosis,” which is often and historically more correctly called “conglomerative” or “chronic nodular silicosis,” the silicotic nodules coalesce into a fibrotic mass and a confluence of opacities on the chest film (“progressive massive fibrosis” is a term more historically correct for coal workers’ pneumoconiosis). This leads to contraction of the upper lobes, traction emphysema, and bullae, which contrast against the densely fibrotic parenchyma and create the “angel wing” pattern (see Figure 17.10). Clinically, this massive fibrotic reaction in the lungs leads to dyspnea and progressive restrictive disease. There may also be a concealed obstructive



**Figure 17.9.** Advanced simple silicosis. Note micronodular opacities, hyperplastic lymph nodes, and calcification of hilar nodes.

component reflecting some underlying emphysema or COPD (chronic obstructive pulmonary disease). Unlike asbestosis, there are few physical signs that accompany even advanced silicosis; rales and clubbing are generally absent. Ultimately, respiratory failure and right heart failure (cor pulmonale) may result. Eventually, the outcome of advanced conglomerative silicosis is respiratory failure and cardiopulmonary arrest if the disease continues to progress. There is no way to modify progression of the disease.



**Figure 17.10.** Conglomerative silicosis in a high-performance ceramics worker; the patient compounded the material, which included silica, for the tiles on the heat shield of the Space Shuttle. (Reproduced with permission of the American Academy of Family Practice.)

“Accelerated silicosis” is a form of silicosis that is rapidly progressive and passes through the simple and complicated stages within a few years or even months. It is associated with heavy dust exposure and carries a very poor prognosis. Accelerated silicosis usually occurs in outbreaks associated with abusive working conditions. This form of silicosis is thought to have been responsible for most of the deaths of the workers in the “Hawk’s Nest” water tunnel disaster of 1930–1931 at Gauley Bridge, West Virginia, the worst industrial disaster in American history. The last reported outbreak in North America came in 1988 in a cluster of cases among oil field sandblasting workers in Texas.

Patients with silicosis who are already infected with the tubercle bacillus, other mycobacteria, or similarly behaving organisms (such as actinomycetes) are very susceptible to opportunistic infection, most commonly reactivation of tuberculosis. The resulting, distinct condition of "silicotuberculosis" can become a devastating, swiftly progressive fibrotic process that, untreated, resembles a malignancy. The diagnosis of silicotuberculosis is suggested when there is a change in the rate of progression of silicotic nodules on the chest film so that the disease appears to be accelerating or when there is onset of systemic symptoms such as fever, weight loss, or new cough (especially hemoptysis). The infection tends to favor upper lobes and may develop early when fibrosis is not far advanced. The recovery of acid-fast bacilli from sputum may be difficult because the mycobacterial burden is much less than in other forms of tuberculosis; repeated sputum induction and even gastric washings may be required. *Mycobacterium tuberculosis* is the most common organism, atypical mycobacteria are becoming increasingly common (*Mycobacterium kansasii*, *Mycobacterium avium-intracellulare*), and a similar condition occurs with opportunistic infection by *Acinetobacter spp.* (See Figure 17.11, which is an advanced case of silicotuberculosis with infection by *M. kansasii*.) Deep fungal infections (aspergillosis, cryptococcosis, sporotrichosis) and nocardiosis have also been described. Silicotuberculosis tends to be refractory to treatment due to poor penetration into the fibrotic mass by antibiotics. Often, patients require life-long suppression using triple therapy.

"Acute silicosis" is a distinct disease, characterized by an alveolitis that occurs early in the disease due to overwhelming silica dust exposure, which produces a clinical picture resembling diffuse alveolar injury. Patients rarely survive to progress to more chronic forms and so the pathology does not feature the typical lesions of silicosis. Acute silicosis, or silica-induced alveolar proteinosis, is associated with massive outpouring of proteinaceous debris and fluid into alveolar spaces, presumably as a consequence of acute inflammation and diffuse alveolar injury. The chest film appears

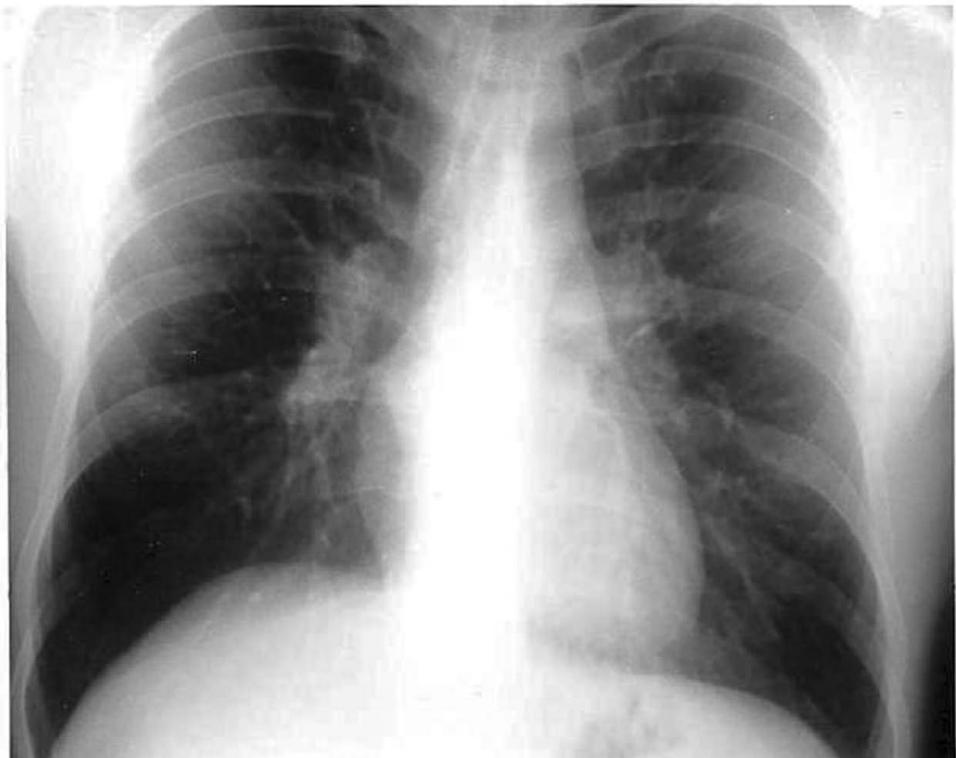


**Figure 17.11.** Silicotuberculosis with an atypical mycobacterium, infection due to *Mycobacterium kansasii*.

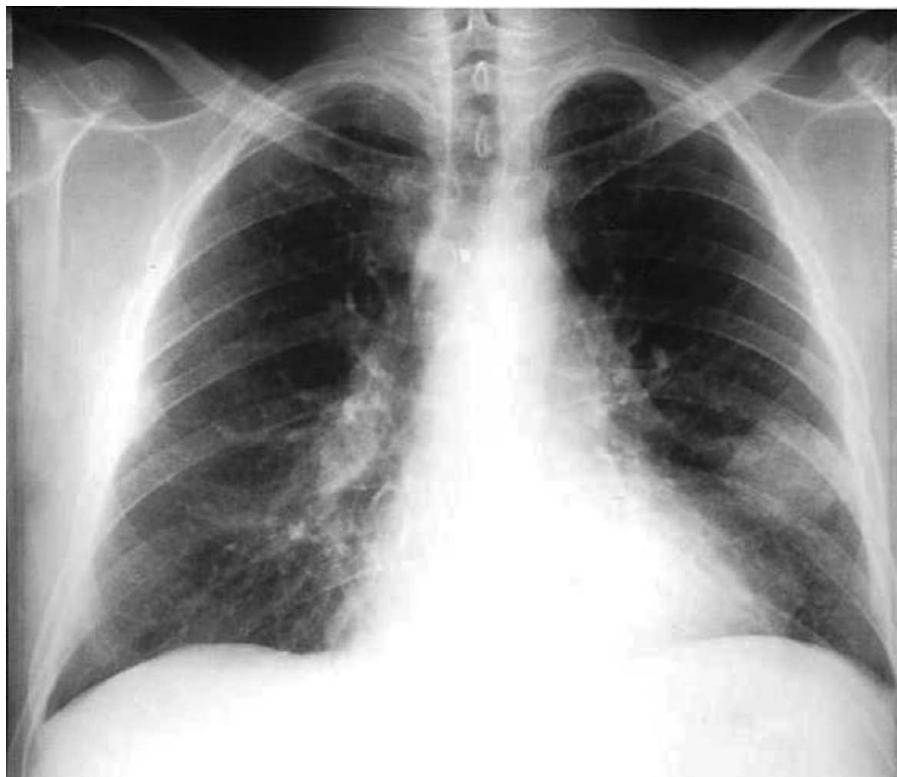
much like a slowly evolving pulmonary edema or adult respiratory distress syndrome (ARDS) but without any obvious precipitating cause. Acute silicosis may respond in part to lavage and administration of high-dose steroids, but is still generally fatal since respiratory failure supervenes due to continued accumulation of debris and lipoproteinaceous material in the airspaces. Occasionally acute silicosis is complicated by opportunistic infection. In view of the extremely poor prognosis and tendency for this form of silicosis to be seen in young workers because of the circumstances of exposure, the physician must consider transplantation and other heroic measures. Like accelerated silicosis, it has been associated with

multiple fatalities among workers who have been heavily exposed at the same site.

Silicosis may be associated with other silica-related disorders, such as systemic sclerosis (including scleroderma, the skin manifestations), the combination of which is called Erasmus' syndrome, and its characteristic nephritis, a more benign nephritis with proteinuria, and airway obstruction, although the pattern of advanced silicosis is always predominantly restrictive. (See Figure 17.12, the chest film of a man with biopsy-proven silicosis, showing mostly hilar changes including calcification of nodes,



**Figure 17.12.** Silicosis, with early eggshell calcification on right, in a patient with systemic sclerosis (Erasmus syndrome) and nephropathy.



**Figure 17.13.** Lung cancer in a worker with advanced simple silicosis and hilar changes.

who had systemic sclerosis and classic scleroderma, with proteinuria and early CRST nephropathy.)

Silica exposure, and therefore these disorders, is also associated with an elevated risk of lung cancer. (See Figure 17.13. A single case does not establish an association, of course, but this case illustrates the presentation of lung cancer against a background of parenchymal and hilar silicosis.) This elevation is associated with significant exposure to crystalline silica but does not seem to require radiological silicosis and is interactive with cigarette smoking, although not to the same degree as asbestos.

### *Asbestosis and Nonmalignant Asbestos-Related Disorders*

This section discusses only nonmalignant respiratory diseases related to asbestos, which are growing less common because there are fewer people living today who had asbestos exposure in the 1970s and before when exposure levels were high. Even so, there are many prevalent cases among older workers and occasional new cases. The Centers for Disease Control and Prevention (CDC) has determined that the total burden of asbestosis in North America is probably at its peak at the current time, because of the latency of the disease, which requires a decade or more to develop, depending on the intensity of exposure.

Exposure to asbestos in a young person today is most likely to occur in poorly managed asbestos abatement activities. Asbestos as a hazard is discussed in Chapter 10; asbestos-related malignancy is discussed in the section on occupational cancer. Although there is an on-going debate about the relative potency of chrysotile compared to the amphiboles in causing these cases, there is no serious controversy over whether chrysotile can cause these disorders. A unified approach to the diagnosis of nonmalignant asbestos-related disease, with standardized criteria, has been developed by the American Thoracic Society and is now in its second revision (see Table 17.2).

“Asbestosis,” sometimes called “white lung” colloquially, is a functionally serious pneumoconiosis, lethal in its advanced forms, resulting from the inhalation of large quantities of asbestos fibers. Asbestosis should never be used as a general term for a history of exposure to asbestos or for other asbestos-related conditions. The term is specific to the interstitial fibrotic lung disease. Even so, asbestosis usually occurs in association with other nonmalignant asbestos-related disorders and is a strong risk factor for malignancy, which may not be easily detectable because of the chest film changes of the underlying pneumoconiosis.

Asbestosis is a widespread interstitial fibrosis that develops first in the lower lung fields as a pattern of “irregular opacities,”

**Table 17.2.** The Approach to Nonmalignant Asbestos-Related Respiratory Disorders

2004 Criterion and Guidelines	How to Satisfy the Criterion
Evidence of structural change, as demonstrated by one or more of the following: • Imaging methods • Histology	Demonstrate the existence of a structural lesion consistent with the effects of asbestos. This criterion is almost always satisfied by chest film in the first instance. Chest film, HRCT, and possibly future methods based on imaging. Criteria for identifying asbestosis on microscopic examination of tissue have been developed by the College of American Pathologists. Biopsy is seldom required to make diagnosis of nonmalignant asbestos-related disease and may be contraindicated in advanced diseases due to risk of pneumothorax. Latency for asbestosis is typically ten to twenty years or more. Exposure was higher in the workplace in the 1970s and earlier.
Evidence of plausible causation, as demonstrated by one or more of the following: • Occupational and environmental history of exposure (with plausible latency) • Markers of exposure • Recovery of asbestos bodies	See Chapters 8 and 15. Bystander and passive exposure may occur. The most likely contemporary exposure for young people is in asbestos abatement activities that are not in compliance with the OSHA Asbestos Standard or other applicable standards. Sufficient to establish asbestos exposure if no other cause of pleural disease demonstrated. The chief marker of exposure is the presence of pleural plaques. Asbestos bodies are asbestos fibers that are coated with protein and hemosiderin as a result of macrophage ingestion. They may be demonstrated through sputum examination, bronchoalveolar lavage, biopsy (rarely required), or autopsy. Individual asbestos fibers are rarely visible by light microscopy during examination of tissue, even when numerous.

Exclusion of alternative diagnoses

Evidence of functional impairment, as demonstrated by one or more of the following:

- Signs and symptoms (including crackles)
- Change in ventilatory function (spirometry)
- Impaired gas exchange (reduced diffusing capacity)
- Inflammation
- Exercise testing

Nonmalignant diseases resembling asbestos-related disease should be ruled out clinically: sarcoidosis, idiopathic interstitial fibrosis, IV drug-injection fibrosis, systemic sclerosis and other rheumatological lung diseases.

Functional assessment is not required for diagnosis but is part of a complete evaluation, in order to monitor progression and to quantify impairment.

Signs and symptoms are not specific for the diagnosis of asbestosis but are associated with significant impairment. The progression or new onset of respiratory symptoms is evidence of diminished reserve and decompensation. Cor pulmonale and right heart failure is a late change and a poor prognostic sign.

Typical pattern is early irritation and trend to low-grade obstructive change and small airway effects, usually still within range of normal function, which are soon overwhelmed in advanced disease by progressive restrictive change.

Low diffusing capacity for carbon monoxide is relatively sensitive and often reduced early in the progression of asbestosis, but it is not specific to asbestosis.

Bronchoalveolar lavage is not necessary for diagnosis and is rarely performed but would show inflammation while asbestosis is active. Histology not required but may be indicative.

Useful for work capacity evaluation and for early detection of impairment.

Source: Adapted from the American Thoracic Society.

which sometimes resemble little threads or lint on the chest film. The characteristic opacities are short, squiggly, or irregular lines that often link together rather than the smooth, discrete,



**Figure 17.14.** Asbestosis at an advanced stage. Note fibrotic changes, irregular pleural border and diaphragm, shaggy heart border, and blunted costophrenic angles, all of which are typical. Source: Ad Hoc Committee [on Update of 1986 Criteria for the Diagnosis of Non-malignant Asbestos-Related Disease: Guidotti TL, Miller A, Christiani D, Wagner G, Balmes J, Harber P, Brodkin CA, Rom W, Hillerdal G, Harbut M, Green FHY]. Diagnosis and initial management of nonmalignant diseases related to asbestos. *Am J Respir Crit Care Med*, 2004;170:691-715. Reproduced with Permission from The American Journal of Respiratory and Critical Care Medicine, American Thoracic Society.

rounded opacities of other pneumoconioses. Over time, these coalesce into fibrotic streaks and bands (see Figure 17.14). The lung becomes progressively scarred and stiff. The natural history of the disease is progression of the restrictive impairment, sometimes to total disability, and a high risk of bronchogenic cancer, whether asbestosis is visible on the chest film or not. Asbestosis usually has many accompanying features on chest film that reflect pleural and pericardial changes, including diaphragmatic calcification (which results from pleural plaques on the diaphragm), blunted costophrenic angle (due to pleural fibrosis), and a “shaggy” (ill-defined) left cardiac border (which results from composite shadows of fibrosis and the pericardium). Examination of individuals with asbestosis may reveal clubbing, crackles, and tachypnea.

The management of asbestosis is supportive and preventive; patients with asbestosis may experience serious and sudden decompensation in the event of intercurrent pneumonia. The disease is not associated with smoking, but smoking clearly makes the symptoms worse and management more difficult. For this reason, and because there is an interaction between smoking and asbestos exposure for cancer risk, quitting smoking is even more critical for such patients than for other smokers.

All but the earliest irregular parenchymal changes are usually accompanied by characteristic pleural signs on the chest film (shown well in Figure 17.14). These result from a transient, painless pleuritis that occurs early in the natural history of asbestos-related disease but is almost never observed radiologically unless captured by accident. Pleural changes alone do not define asbestosis, nor do they indicate a risk of mesothelioma beyond confirming the fact of asbestos exposure. “Pleural plaques” are discrete fibrotic regions of the parietal pleura visible on the chest film as an irregular pleural border from the side and as polygonal lesions overlying the lung fields face on. Pleural plaques tend to calcify over time. On the diaphragm they appear as rectangular-stepped discontinuities or streaks of calcification. They can be very large and dramatic (as in



**Figure 17.15.** Pleural plaques, most prominently on right postero-lateral chest wall, in a worker exposed to asbestos. Source: Ad Hoc Committee [on Update of 1986 Criteria for the Diagnosis of Nonmalignant Asbestos-Related Disease: Guidotti TL, Miller A, Christiani D, Wagner G, Balmes J, Harber P, Brodkin CA, Rom W, Hillerdal G, Harbut M, Green FHY]. Diagnosis and initial management of nonmalignant diseases related to asbestos. *Am J Respir Crit Care Med*, 2004;170:691-715. Reproduced with Permission from The American Journal of Respiratory and Critical Care Medicine, American Thoracic Society.

Figure 17.15) or tiny and ill-defined, bilateral or unilateral, overlying a single rib (the usual location), or extending across several intercostal spaces. Pleural plaques often occur in isolation and serve as reliable markers (in the absence a history of chest trauma or tuberculosis) of past exposure to asbestos. “Pleural thickening” is a more diffuse fibrotic change in the pleura, without a discrete border. Pleural changes due to asbestos almost never cause symptoms or functional impairment.

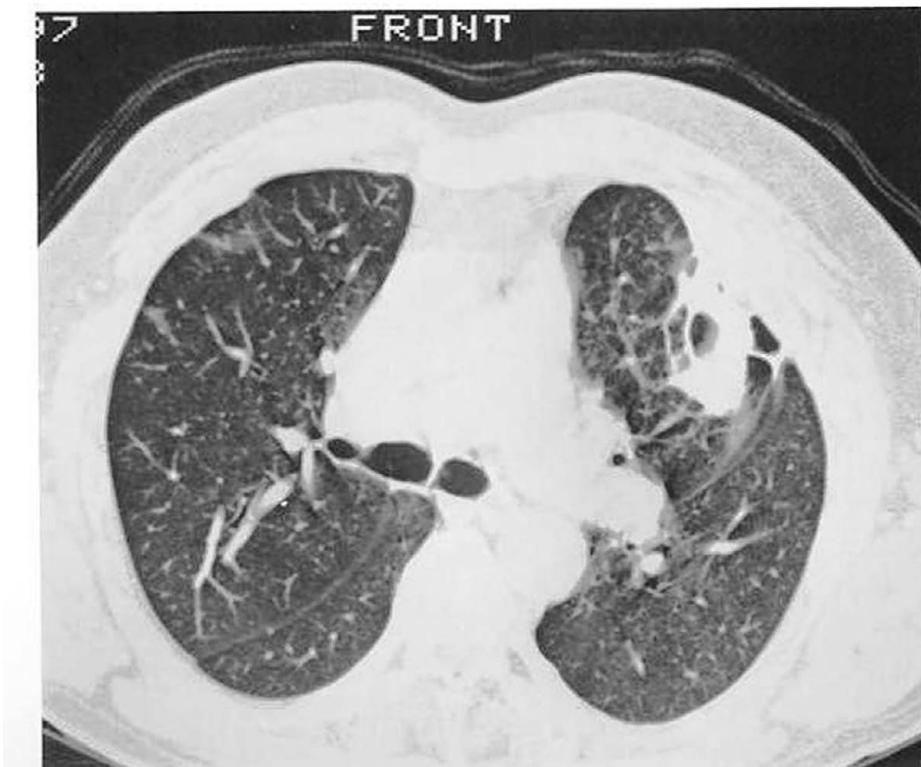
Less common effects of asbestos exposure on the pleura include benign pleural effusion (the transient pleuritis described above)

and rounded lung syndrome. Benign pleural effusion does not appear to be a premalignant lesion and behaves much like the effusions associated with primary tuberculosis. “Rounded atelectasis” or “folded lung” presents as a round, comet-shaped pleural-based mass with the “tail” pointed toward the hilum. (Figure 17.16 shows this, in the left mid-lung field.) This mass forms when a peripheral segment of lung becomes atelectatic and due to the traction of fibrosis rolls up on itself, caught between thickened visceral pleura and interlobar fissure. Rounded atelectasis is often mistaken for lung cancer but has a characteristic appearance on HRCT, with curved blood vessels, air trapping, and attachment to the pleura, as in Figure 17.17.

Asbestos bodies (asbestos fibers encrusted with protein and iron from cellular activity) are usually abundantly visible when they are



**Figure 17.16.** Rounded atelectasis, a peripheral mass consisting of parenchyma distorted by pleural fibrosis; this is not a lung cancer. Source: Ad Hoc Committee [on Update of 1986 Criteria for the Diagnosis of Nonmalignant Asbestos-Related Disease: Guidotti TL, Miller A, Christiani D, Wagner G, Balmes J, Harber P, Brodkin CA, Rom W, Hillerdal G, Harbut M, Green FHY]. Diagnosis and initial management of nonmalignant diseases related to asbestos. *Am J Respir Crit Care Med*, 2004;170:691–715. Reproduced with Permission from The American Journal of Respiratory and Critical Care Medicine, American Thoracic Society.



**Figure 17.17.** CT scan of same patient as in Figure 17.16, showing details of the mass: air trapping, contorted blood vessels, attachment to pleura. Source: Ad Hoc Committee [on Update of 1986 Criteria for the Diagnosis of Nonmalignant Asbestos-Related Disease: Guidotti TL, Miller A, Christiani D, Wagner G, Balmes J, Harber P, Brodkin CA, Rom W, Hillerdal G, Harbut M, Green FHY]. Diagnosis and initial management of nonmalignant diseases related to asbestos. *Am J Respir Crit Care Med*, 2004;170:691-715. Reproduced with Permission from The American Journal of Respiratory and Critical Care Medicine, American Thoracic Society.

searched for under the microscope (Figure 17.18). They are also markers of exposure to asbestos.

Airways are also affected by asbestos exposure, as a result of interstitial fibrosis in the adjacent alveoli. Chronic obstructive airway disease caused by asbestos may be clinically significant in some cases and adds an obstructive element to the primarily restrictive impairment of asbestosis. Early on, there may be a relatively pure detectable obstructive component on spirometry, but as the disease progresses the restrictive pattern soon becomes overwhelmingly predominant.



**Figure 17.18.** Asbestos body visible in lung tissue on a scanning electron microphotograph. (Photograph courtesy of Dr. Jerrold L. Abraham.)

### ***Coal Workers' Pneumoconiosis and Less Fibrogenic Dusts***

“Coal workers’ pneumoconiosis” (CWP), or “black lung,” is probably the best-known dust disease of the lung because of historical associations with mining communities. The characterization of CWP as a distinct pneumoconiosis came slowly because the historically recognized lung disease of coal miners (“miner’s phthisis”) was often confounded in epidemiological studies and complicated in individual

cases by silicosis, cigarette smoking, and a high prevalence of tuberculosis and histoplasmosis in coal mining regions in North America, especially Appalachia. It has been recognized for decades, however, that CWP is a distinct disease with a tendency to develop into a more aggressive fibrotic form; the functional impairment in nonsmoking miners is obstructive and is initially somewhat less than for other fibrogenic dusts but can progress into severe restrictive disease.

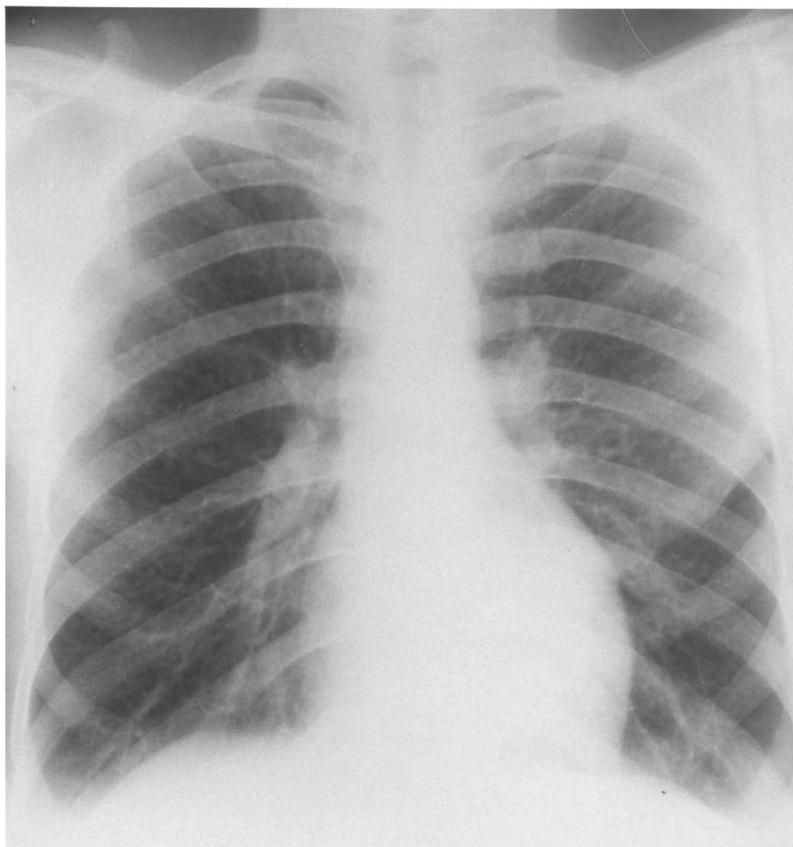
CWP and other lung diseases associated with underground coal mining are declining in frequency as a result of dust suppression in mines but have not been eliminated in the United States. CWP also remains an important and common occupational lung disease in developing countries.

The medical diagnosis of CWP almost always depends on the chest film. Only rarely is biopsy justified but the diagnosis may be made by direct examination of tissue at autopsy to qualify the surviving family for benefits. As noted, CWP has its characteristic histological lesion, the coal dust macule, a fibrotic nodule, less dense than the silicotic nodule, surrounding an arteriole and packed with black "anthracotic" particles left behind by macrophages that migrated to that location.

The chest film in CWP begins, as do most pneumoconioses, with scattered rounded opacities, predominantly in the lower lung fields initially. Figure 17.19 demonstrates the typical chest film appearance of CWP. The interpretation of the chest film in CWP relies on the occupational history, since there is no pathognomonic sign for early CWP. However, the natural history of CWP is distinct from that of other pneumoconioses. CWP progresses through stages, first as "simple CWP," in which the opacities remain discrete and lung function is preserved, and then, when any one mass reaches one centimeter in diameter, to "complicated" CWP, in which they coalesce into complex, structured disc-like lesion that appears in the upper lung fields, and restrictive changes become apparent. As these lesions grow by accretion of smaller opacities, they may form very large, much denser masses. This is called "progressive massive fibrosis" and is associated with restrictive

impairment, pulmonary hypertension, a high risk of right heart failure, and a poor prognosis. Complicated CWP has disappeared in North America but was once all too common where coal was mined underground.

It is not clear why CWP should project a radiological shadow in the first place, since coal is mostly elemental carbon. A density of 0.4 g/cm<sup>3</sup> is required for coal dust to be barely visible on the chest film, and the total lung burden of dust in the lung is far lower than what



**Figure 17.19.** Simple coal workers' pneumoconiosis in a 55-year old retired Japanese anthracite miner. Reprinted with permission from "Occupational Lung Diseases," February, 1984, *American Family Physician*. Copyright © 1984 American Academy of Family Physicians. All Rights Reserved.

would be required to produce shadows by density of dust alone. This strongly implies that the origin of the "interstitial" pattern of small nodular opacities characteristic of CWP is local tissue reaction.

There is no practical method of detecting or quantifying coal dust in the lung of the living patient other than by x-ray film. Sputum production of coal pigment is compatible with a recent acute exposure but says nothing about the cumulative lung burden.

The potency of coal dust in inducing changes in the lungs is reflected in the "rank" of the coal, which is its silica content and which is a complex function of the geologic characteristics of the coal seam. Higher-rank coal, such as anthracite, is more fibrogenic than bituminous coal, which in turn is more fibrogenic than lignite (brown coal). The rank and density of coal reflects the compaction and mineralization of fossil wood and plant matter over millennia.

The pathology of CWP involves the bronchiole, the arteriole, and the alveolar wall. The characteristic pathologic finding in CWP is the coal dust macule, a discrete (occasionally confluent) nonindurated focus of black pigment macroscopically under the visceral pleura and on the cut surface of lung parenchyma. The macules tend to constrict and efface arterioles that run through them. Microscopically, the coal dust macules are typically in various stages of development. The carbonaceous pigment is mostly intracellular in macrophages and may remain present ten years after cessation of exposure. The macrophages are concentrated in the peribronchial and periarterial interstitium, trapped by local fibrosis. The distribution of macules follows the known patterns of peribronchial and periarterial lymphatic drainage, and draining nodes typically show heavy accumulations of pigment. Limited dilatation of the respiratory bronchioles is common, in a pattern compatible with centrilobular emphysema, but without bronchiolitis.

With so many structures involved, one might expect that simple CWP would have a profound effect on ventilation, perfusion, and gas exchange. Surprisingly, it usually does not. In general, coal miners have slightly lower forced vital capacities (FVC) and one-second

forced expiratory volumes ( $FEV_1$ ) than do nonminers; these changes usually do not become statistically significant until the miners have worked underground thirty years or more. The FVC and  $FEV_1$  are lowest, on the average, for anthracite miners and less reduced for bituminous miners but are not markedly reduced in any large population sample of working miners. Cigarette smoking obliterates any differences in the degree of small airway disease associated with simple CWP. Diffusing capacity is within normal limits for nonsmoking miners with simple CWP. When the disease converts to progressive massive fibrosis, however, the advancing restrictive process comes to overwhelm the earlier, more modest changes.

“Caplan’s syndrome” is the association of CWP with rheumatoid arthritis; it can occur with other pneumoconioses. Caplan’s syndrome probably represents one particularly well-defined end of the spectrum of autoimmune disorders caused by altered self-antigens in the damaged lung. Anti-lung antibodies and other manifestations of autoimmunity are common in the pneumoconioses generally, particularly CWP.

### Toxic Inhalation

Toxic inhalation is a general term for a pattern of deep-lung injury produced by a variety of gases that penetrate to the deep lung, including ozone ( $O_3$ ), phosgene ( $COCl_2$ ), chlorine ( $Cl_2$ ), nitrogen oxide ( $NO_2$ , not to be confused with nitrous oxide,  $N_2O$ , or nitric oxide, NO), hydrogen fluoride (HF) (see Table 17.3). Exposure to these gases at the levels required to produce this condition is usually the result of accidental release, uncontrolled chemical reactions, or fires. Some of these gases, particularly phosgene, chlorine, and nitrogen dioxide, are generated when plastic furnishings and interior design fixtures burn, as in a hotel fire. In such combustion situations, cyanide (CN) and carbon monoxide (CO) are also released and may contribute to toxicity. Several gases uncommonly encountered in industry and agriculture may cause toxic inhalation, with other systemic toxicity: silane ( $SiH_4$ ), phosphine ( $PH_3$ ), hexafluoroacetone

**Table 17.3.** Some Toxic Gases Associated with Diffuse Alveolar Injury

---

Oxidant gases (nitrogen dioxide, ozone, phosgene, hyperbaric oxygen)

Sulfur dioxide (rare because of low penetration to deep lung)

Chlorine

Ammonia

Hydrogen sulfide

Hydrogen fluoride

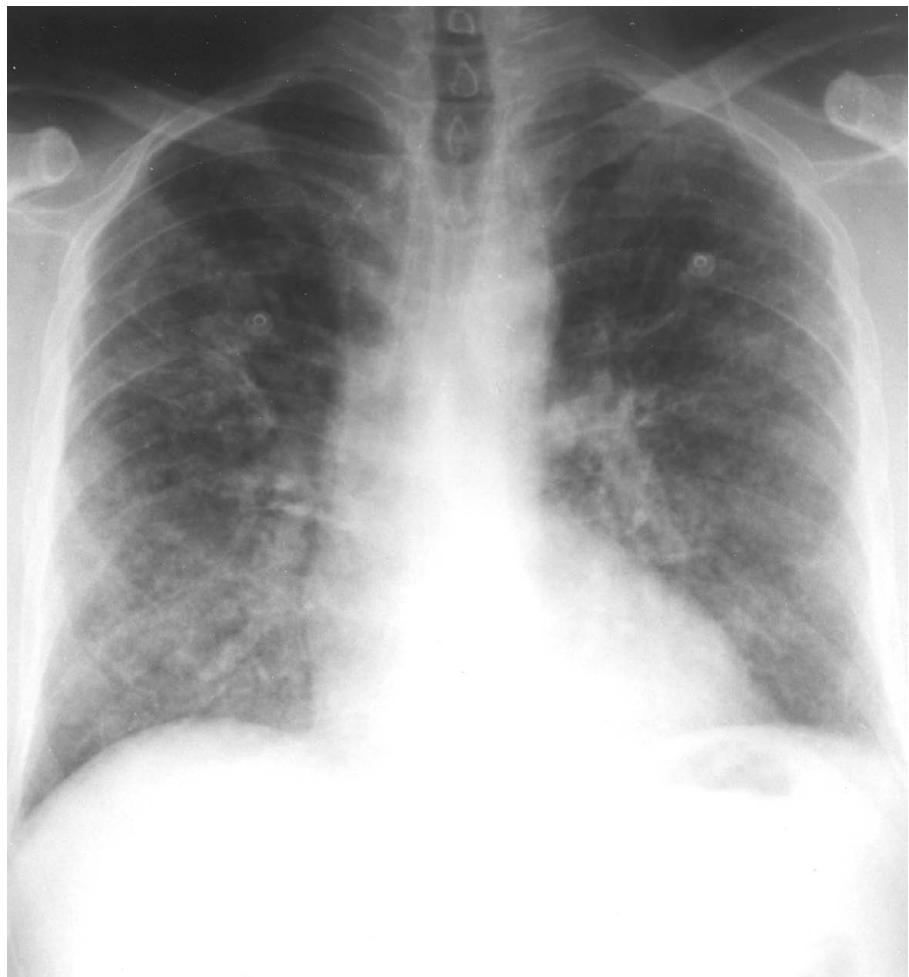
Heated mercury, cadmium (not to be confused with metal fume fever)

---

( $\text{F}_3\text{CCOCF}_3$ , found in the semiconductor industry), and methyl bromide ( $\text{CH}_3\text{Br}$ , bromomethane) and sulfuryl fluoride ( $\text{SO}_2\text{F}_2$ ), used as fumigants. Nickel carbonyl is discussed at the end of the subsection because of its unusual characteristics.

Historically, the most common toxic inhalation was silo filler's disease, which resulted from exposure to nitrogen dioxide released during the incomplete fermenting process during storage in a large, deep silo (see Figure 17.20). When agricultural workers entered the silo a few weeks after filling it with harvested grain (especially maize, because it is rich in tryptophan, a nitrogen source), they would encounter the brownish gas and experience the characteristic symptoms described. Greater awareness and changes in agricultural practices have mostly eliminated the disorder. (Silo filler's disease is not to be confused with farmer's lung, a hypersensitivity pneumonitis.)

The mechanism of toxic inhalation involves inhalation of a gas that is either not readily dissolved in water, and so penetrates to the deep lung at a high concentration, or inhalation of a soluble gas at such a high exposure that it reaches a high concentration at the alveolar level regardless. In addition to incidental irritation of the upper airway (and usually eyes), the gas irritates the bronchial mucosa causing severe cough and may induce acute bronchospasm and damages to the bronchiolar mucosa on the way down. Once the gas has penetrated to the alveolar level, it injures the delicate alveolar wall, damages the barrier between the vascular space and first the interstitium and then



**Figure 17.20.** Silo filler's disease, toxic inhalation due to nitrogen dioxide, with infiltrates representing pulmonary edema. Reprinted with permission from "Occupational Lung Diseases," February, 1984, *American Family Physician*. Copyright © 1984 American Academy of Family Physicians. All Rights Reserved.

the alveolar space, and may cause pulmonary edema. This "diffuse alveolar damage" results in a relatively slow process by which fluid passes under vascular pressure from the alveolar capillary bed into the alveolar wall as interstitial edema and then later breaks through under pressure into the alveolar space as pulmonary edema. If the patient

recovers, the diffuse alveolar injury may be repaired in a dysfunctional manner, with scarring and interstitial fibrosis. Injury to small airways may also result in overly exuberant fibrosis that overgrows the wall of the airway, leading to bronchiolitis obliterans (not to be confused with bronchiolitis obliterans organizing pneumonia, which is different), in which small airways disappear because they are absorbed into the inflammation around them, which leads to fixed airway obstruction and chronic impairment.

As a consequence of this sequence of events, these agents produce a relatively stereotyped clinical picture, with a predictable but not always complete sequence of events:

1. Acute bronchospasm and dyspnea develop at the time of exposure. Usually the subject is removed from the exposure by co-workers, who may also be affected if they do not stop to use self-contained breathing apparatus.
2. Unless overwhelmed, the victim will often appear to recover from the initial symptoms.
3. About four to twelve hours later, severe dyspnea may develop as a result of pulmonary edema caused by diffuse alveolar injury that may not have been appreciated at the time of initial exposure. The pattern of pulmonary edema on chest film is similar to that of the adult respiratory distress syndrome, or “capillary leak,” rather than pulmonary edema due to congestive heart failure. The time lag reflects the migration of fluid from the vascular to the alveolar space and represents the onset of alveolar pulmonary edema.
4. After the patient recovers from the acute process and the delayed development of pulmonary edema, their recovery may be interrupted by the onset of refractory shortness of breath weeks later. This would be due to bronchiolitis obliterans. A chest film may show a hyperlucent lung and air trapping.
5. Much later, interstitial fibrosis may occur, with the characteristic pattern of “honeycombing” (interstitial fibrosis) on the chest film.

The management of patients with toxic inhalation is primarily supportive. This form of pulmonary edema is similar to adult respiratory distress syndrome (ARDS) and is familiar in the intensive care unit. One-third of such patients die.

The single greatest clinical mistake in such cases is not to recognize the potential for pulmonary edema. Patients have been sent home from the emergency room only to die in their bed hours later. Pulmonary edema due to toxic inhalation is life threatening and leads to mortality in about one-quarter of cases historically. Among those who survive the pulmonary edema, an unknown fraction will then develop the obliterative bronchiolitis. A few develop bronchiolitis without experiencing pulmonary edema. There is no effective treatment for obliterative bronchiolitis after it is fixed. However, as it develops the process may be arrested with steroids, which are sometimes required for prolonged periods. The small airways or bronchioles have in the aggregate a much greater cross-sectional area than the larger airways associated with asthma. There are, of course, many more of them. As a result, their obstruction does not interfere significantly with airflow until very many are involved. This means that the disorder is largely silent until the condition is far advanced. Symptoms of cough and shortness of breath may not be associated with the degree of the patient's impairment in pulmonary function. These symptoms relate more closely to the fibrotic changes occurring in the tissue than the obstruction of the airways.

Two precautions should be followed in every suspected case of toxic inhalation. First, the patient must be observed through the period of risk for pulmonary edema, normally in a hospital. At the very least, a responsible adult must stay with the patient in order to return the patient to medical attention immediately if dyspnea develops. Second, exposure to other toxic agents must also be considered, particularly in cases of fire or explosion. The most important gases in such multiple exposures are carbon monoxide and hydrogen cyanide, but pesticides, hydrogen sulfide, volatile organic compounds that depress the central nervous system, or other chemicals may be present, depending on the circumstances of the exposure.

An organic metal compound, nickel carbonyl ( $\text{Ni}(\text{CO})_4$ ), causes an unusually treatable form of toxic inhalation, with elements of systemic toxicity. Nickel carbonyl produces a similar pattern of pulmonary edema and is often associated with fever, leukocytosis, and headache; it is most often encountered in nickel plating and case-hardening operations in treating steel and in nickel refining. Urinary nickel levels confirm the diagnosis. Specific treatment with disulfiram, parenterally if necessary, may be life-saving by reducing biologically available nickel following exposure. This condition is quite rare.

### **Occupational Asthma**

Asthma is, of course, very familiar to any clinician. However, occupational and environmental asthma require new ways of thinking about this disorder and an expanded definition of asthma. Usually asthma of whatever type presents as wheezing, cough, and shortness of breath, occurring repeatedly in isolated episodes, often immediately following exposure to a recognizable allergen. However, there are many variations. Airflow obstruction may become chronic (although still variable), and a restrictive defect may develop as a result of air trapping. Asthma may also present as a bronchitis, without discrete episodes of bronchospasm.

Asthma can be considered to be a disease with many variations, which has been the predominant view in clinical medicine, or a collection of individual disorders all characterized by reversible obstruction of air flow, latent or active airway reactivity, and some degree of inflammation of the airways. Occupational and environmental asthma is most usefully thought of as an umbrella term for a group of related disorders involving the response of airways to triggers and conditioning factors in the work environment. Intensive study of occupational asthma, in particular, in recent years has expanded medical knowledge of asthma in general and its mechanisms.

“Environmental” asthma hardly needs to be qualified as such because environmental factors are so well known to play a major role in triggering asthma from any cause: environmental allergens (most often

molds, dust mite, cockroach, and animal antigens), air pollution (particularly ozone, as a trigger for asthmatic episodes), dampness (a known independent risk factor for respiratory symptoms in homes), chemical irritants (passive cigarette smoke, strong fragrances especially those rich in aldehydes, volatile solvents, and gasoline), dust (acting as a nonspecific irritant), and cold, dry air (which provokes cold-induced and exercise-induced asthma). This does not mean that asthma is caused exclusively by environmental exposures in the first instance, however. Asthma is clearly a complex process that begins with a genetic predisposition, is conditioned by environmental factors (which may include air pollution), is shaped in its expression by either sensitization (to allergens in the surrounding environment, acting by a very specific hypersensitivity response) or irritants (acting by a nonspecific mechanism), and triggered or perpetuated by environmental exposures acting on a persistent underlying condition of inflammation and airway reactivity. Whether the case in question is "extrinsic" asthma (the classic immediate hypersensitivity variety typically seen in childhood) or "intrinsic" (the more variable variety occurring in adulthood), asthma is so closely linked with environmental influences that it practically defines an environmental disorder.

"Occupational asthma" as a term of art refers to reversible airway disorders arising from or triggered by factors in the workplace and embraces a broad range of symptoms, including the classic immediate hypersensitivity reaction of allergic asthma, immediate bronchospasm due to airway provocation, an isolated late response that sometimes presents as sleep disorder, and combinations and variations of these presentations. A wider variety of airway responses than are typically seen in conventional asthma is common to occupational asthma. As well, the spectrum of incomplete or partial responses is more complete because symptoms that characterize asthma (especially cough and shortness of breath) are commonly seen in workers with underlying reactive airways but who have not been given the formal diagnosis of asthma, because the disorder is predominantly upper airway and may not present with wheezing. These patients may have allergic rhinitis or allergic sinusitis and are usually atopic. Sometimes they had

asthma in childhood and think that they have “outgrown” it. (As children grow into adulthood, the increase in diameter of their airways reduces obstruction to airflow but airway reactivity remains.)

Occupational asthma is easiest to understand in terms of its relationship to exposures in the workplace. One of the most important practical distinctions to be made is between conditions induced by a sensitizer (antigens) and those induced by irritants. The former is more difficult to manage because a sensitized individual experiences an immune response after exposure at very low levels, well below occupational health standards, and personal protective equipment is often insufficient to control symptoms, usually leading to medical removal that may involve disruptive reassessments, lost jobs and opportunities, and lost income. The latter usually involves moderate to heavy exposure to only moderately irritating gases, solvents, or aerosols (highly irritating inhalants cause toxic inhalation), can often be managed by ventilation of personal protective equipment, and only requires medical removal as a last resort.

Occupational asthma includes the following airway disorders as they relate to workplace exposures:

- New-onset asthma
  - Sensitizer-induced (immediate or delayed response)
  - Irritant-induced
- Provocation and aggravation of preexisting asthma and airway reactivity
- Reactive airway dysfunction syndrome (RADS)
- Cold- or exercise-induced asthma
- Airway reactivity secondary to hypersensitivity pneumonitis

Cold air- and exercise-induced asthma are the same phenomenon, provoking the airway response due to dry air passing over the moist bronchial mucosa, either because of low moisture content or hyperventilation and exertion. This causes drying and cooling of the airway and stimulates vagal receptors, which release mediators from

mast cells for an otherwise conventional asthmatic response of bronchospasm. Cold air and exercise provoke an immediate response, but once exposure or exertion stop, air flow rapidly corrects.

Airway reactivity associated with hypersensitivity pneumonitis is discussed above in the subsection on pneumonitides.

Principles of evaluation in occupational asthma involve a detailed work history that documents the onset of respiratory symptoms and any association with work activity. Symptom diaries are very useful for this purpose. Spirometry establishes levels of function, but because asthma by definition is variable, any one tracing cannot accurately determine baseline impairment. Methacholine challenge is rarely necessary to confirm that airway reactivity exists but is occasionally necessary to ensure that it will be recognized for insurance or compensation purposes. Pre- and post-shift pulmonary function studies may demonstrate a big incremental loss in airflow and may also confirm the diagnosis but do not identify the responsible exposure.

Treatment for all forms of occupational asthma is conventional, as for other forms of reactive airway disease. Avoiding known sensitizers completely and reducing or eliminating any and all irritant exposure is the mainstay of prevention and maintenance. Personal protective equipment may be effective and may allow the worker to continue around irritant exposures but is usually insufficient to prevent responses from sensitizer-induced occupational asthma.

The management of compensation for occupational asthma can be difficult. It is often difficult to support a claim without identification and documentation of a particular trigger (either sensitizer or irritant), although it may not be possible to be that specific in the workplace. Because asthma is by definition highly variable and pulmonary function may be normal between events, impairment ratings tend to be low. The assumption is usually made that because the disorder is episodic and depends on triggers, it does not have to be disabling. This is not necessarily true, if the inciting exposure is technically necessary for a particular job or career that represents the worker's usual occupation or that requires substantial training or

education. There may be substantial dislocations in the injured worker's functional capacity that are not reflected in the ultimate impairment rating and in his or her life and career path that are not reflected in the resulting disability ratings.

### ***Sensitizer-Induced Occupational Asthma***

Sensitizer-induced occupational asthma involves an allergic response to a specific antigen, either a low-molecular-weight hapten (such as an isocyanate or trimellitic anhydride) that binds to a carrier molecule (such as albumin) to form a complete antigen or a high-molecular-weight molecule, often a protein (as in the case of baker's asthma). Sensitization can occur at any level of exposure but is more likely with high exposure levels. Thereafter, however, low levels are sufficient to trigger a hypersensitivity response. Sensitizer-induced asthma may present as an immediate hypersensitivity reaction, when the response is mediated by reaginic antibody, mostly IgE. It may also present as a delayed, leukotriene-mediated response that occurs hours after exposure. Individuals with atopy are more likely to become sensitized on or off the workplace, but a personal or family history of atopy or allergies is not sufficiently predictive of future risk to exclude an otherwise fit worker from a job involving exposure to known sensitizers.

Patterns of airway reactivity may be complicated and mixed in occupational asthma. As noted, at least two types of reactions may occur, singly or in combination. The immediate hypersensitivity reaction is a typical acute asthmatic episode. The delayed hypersensitivity reaction may occur several hours after exposure, consists of dyspnea, cough, and sometimes wheezing, and often interferes with sleep. The delayed reaction may be overlooked entirely in the patient's evaluation or may be falsely attributed to an antigen present in the home. The delayed reaction is thought to be mediated by leukotrienes, a class of mediators of bronchial smooth muscle contraction different from those responsible for the immediate reaction.

Management depends critically on identifying the sensitizing agent in the workplace. Table 17.4 lists common sensitizers. This is by no means a complete list and only hints at the diversity of sensitizing agents. The easiest agents to identify are those that trigger the familiar immediate hypersensitivity reaction soon after exposure. Such conventional allergic sensitizers include animal secretions, ethylene diamine, grain dusts, detergent enzymes (now encapsulated and no longer likely to be a problem), epoxy resin curing agents, and virtually any organic or low-molecular-weight compounds, including metals such as platinum salts.

A few sensitizers produce reactions by mechanisms that are not typical of the common immediate hypersensitivity reaction, such as grain dust, wood dust, formaldehyde, pharmaceutical agents, and the isocyanates (most potently toluene diisocyanate, which has been

**Table 17.4.** Examples of Common Sensitizing Agents in Occupational Asthma

*Low molecular weight*

Isocyanates (entire class; most potent is toluene diisocyanate, abbreviated TDI)

Trimellitic anhydride

Metal salts

Epoxy resins

Fluxes (used in cleaning joints for welding, brazing; historically colophony, a resin, was most potent)

Persulfate

Aldehydes

*High molecular weight*

Pharmaceuticals

Animal proteins (laboratory historically animal dander most potent)

Latex (cross-reacts with avocado, banana proteins)

Cereals

Seafood (crab antigen most potent)

Proteolytic enzymes (historically, when used as additives in detergents)

Wood dust constituents (hardwoods generally more potent)

withdrawn for most purposes), a particularly potently sensitizing family of chemicals used in the production of polyurethane plastics and many paints and coatings. In such cases, the responses may be mixed, with immune, irritant, and toxicological (direct-acting) mechanisms each playing some role. These variant reactions are induced by minute quantities of the antigen, in contrast to those induced by irritant gases or the reactions that result in the hypersensitivity pneumonitides. Dual and variable responses are usually associated with slower recovery times and are relatively refractory to conventional treatment of asthma.

Certain agents tend to produce unusual airway reactions. Exposure to cotton dusts in textile milling and weaving may result in byssinosis ("brown lung"), a slowly developing and sustained reaction lasting several hours and variably associated with chronic changes. Western red cedar dust, a particular problem in the sawmills of the Pacific Northwest and British Columbia, may cause episodic bronchospasm in a cyclic pattern recurring over several days. Grain dust asthma is a complicated response that includes a persistent inflammatory condition and possibly variable responses to different antigens within the dust and its contaminants. Laboratory animal asthma is a particularly severe and rather common form of occupational asthma that carries an unacceptably high risk of systemic anaphylaxis; it occurs among animal handlers in research laboratories and is incompatible with continued employment once it develops. Latex allergy is often expressed as occupational asthma and has become a critical management issue in the healthcare sector, affecting the lives and careers of many nurses.

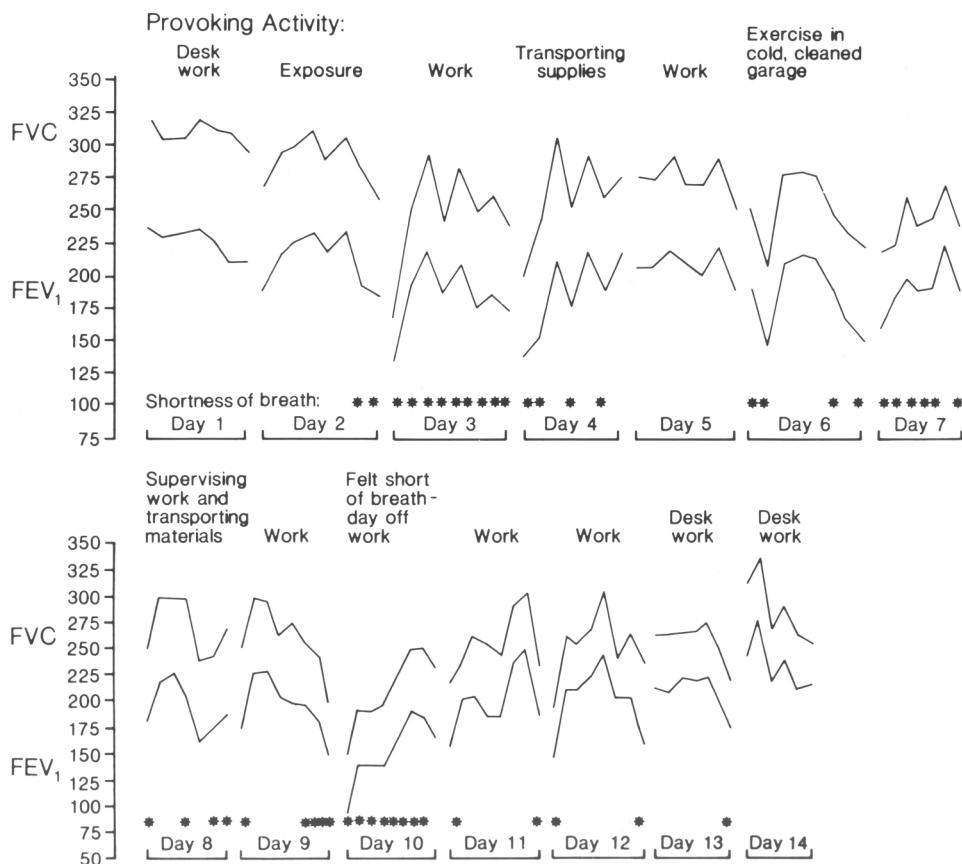
The history alone may be diagnostic in immediate hypersensitivity reactions induced by a single agent used only intermittently in the workplace. Sensitizer-induced occupational asthma can be difficult to evaluate in delayed, dual, or mixed responses. Allergy tests, including skin tests, RAST (radioallergosorbent), and patch tests, are usually helpful in such cases but are often limited in their application in occupational asthma because standard preparations of the suspect chemicals may not be available. Allergic precipitins, the demonstration

in the serum of the patient of precipitating antibody to a particular antigen, are only helpful in suspected cases of hypersensitivity pneumonitis, not occupational asthma, and are developed by many workers who never develop symptoms.

Confirmation of the diagnosis can in theory be obtained by bronchial challenge in cases of immediate and late hypersensitivity reactions but test chambers and laboratories conducting antigen-specific bronchoprovocation are scarce. This is not a commercially viable test because it requires a specialist working in a dedicated, controlled clinical setting using titrated, quantified doses of the suspected antigen aerosolized for inhalation by the patient, following a very meticulous protocol that often requires several hours on successive days. Unless the test is undertaken for research purposes, there are not enough patients to justify the cost of setting it up. There is also a question of liability if the subject has an adverse reaction or is inadvertently sensitized by the test.

A useful substitute test to confirm the diagnosis and identify the agent, one that can be performed in the office, is spirometry before and after the patient reproduces his or her workplace exposure by performing the same job using the same materials. (This is presumed acceptably safe if the worker has already done it as a daily routine.) Detection of the late reaction is especially difficult and may require observing the patient in the office for many hours. Spirometry before and after the work shift may also suggest occupational asthma by reductions in air flow, but small changes in pulmonary function should not be overinterpreted. There is a normal diurnal variation in the peak flow of about 10 percent, more in persons with asthma (in addition to the usual variability in air flow from other triggers), such that lower readings in the afternoon are to be expected. Another useful technique is to provide the patient with a portable peak flow meter and to have the patient record his work duties, symptoms, and peak flow measurements every two hours for two weeks. The resulting patterns can be very revealing. Figure 17.21 shows the pattern in a thirty-seven-year-old tile setter whose asthma was triggered by exposure to the tile cement and grouting.

Once the offending agent is identified, it may be possible to avoid exposure through sound industrial hygiene measures or by changing work practices. More often, the worker must be reassigned to other duties that do not involve exposure to the offending agent, particularly when the mechanism is immunologic.



**Figure 17.21.** Occupational asthma in a thirty-seven-year-old tile and terrazzo setter, whose airway response followed exposure to tile cement or grouting. This graph shows pulmonary function at two-hour intervals during the waking hours over a two-week period in winter. Reaction is not seen on all days, but marked drop in lung function clearly seen following work on several occasions and in response to cold air.

### ***Irritant-Induced Occupational Asthma***

Irritant-induced occupational asthma is more common than sensitizer-induced occupational asthma and may be caused by any irritating exposure. This type of occupational asthma can best be visualized as a continuum from “irritant-induced aggravation of preexisting asthma” (resulting in bronchospasm acutely or perpetuating inflammation chronically) to “irritant-induced new-onset asthma” (in which the irritation induces the airway reactivity in the first instance), and “reactive airway dysfunction syndrome” (RADS), a more severe condition in which airway reactivity arises de novo from airway injury. These distinctions are very important in terms of management, prognosis, and compensation eligibility and management, but there is some degree of overlap.

A well-known example of irritant-induced new onset asthma was called “hot wire asthma” or “meat wrapper’s asthma.” It occurred when meat was wrapped in plastic wrap locally in supermarkets. The early apparatus used a hot wire to cut through the wrap, but the temperature was set too high. The plastic was degraded by the high temperature, releasing irritating fumes. Readjusting the temperature to melt through the wrap, rather than disintegrate it, solved the problem with no inconvenience or loss of productivity.

Irritant-induced aggravation of preexisting asthma (“aggravational asthma” for short) is very common and entirely nonspecific. This is a condition experienced by almost any worker with asthma, hay fever and allergic rhinitis, bronchitis, or for that matter, other airway disorders such as emphysema, who encounters a moderately irritating exposure in the workplace. The irritant exposure makes the existing symptoms of asthma worse in two ways, initially by provoking airway response and over the longer term by adding to the inflammation. Aggravational asthma is usually self-limited, and the worker returns to baseline after avoiding further irritant exposure, which may take a few days. Any irritant exposure will aggravate reactive airways in these circumstances, and all irritating exposures should be removed or reduced, including and especially at home. In a very few cases in which the worker’s underlying

asthma is especially brittle and difficult to manage, aggravation may result in an exacerbation or functional decompensation. Recovery may be prolonged or may only reach a new, lower baseline. The great majority of cases are self-limited.

Irritant-induced occupational asthma of all types is more likely to be associated with a longer term or repetitive exposure to a moderately irritating gas, dust, vapor, or solvent than a brief exposure to a strong irritant, which is more likely to be associated with toxic inhalation. The irritants that induce occupational asthma are much less reactive and are tolerated in the short term. They act by lowering the threshold for bronchial reactivity because of irritation and subsequent inflammation in the airway, sometimes resulting in a sustained nonspecific bronchitis and at other times airway reactivity when provoked by an irritant trigger. Individual susceptibility varies greatly, but the reactions induced by the irritant effects of these agents are generally milder than in sensitizer-induced asthma (with the exception of RADS) and require heavier exposure than those mediated by immunologic processes. Confusingly, some chemicals are both sensitizers (at low concentrations) and irritants (at high concentrations), acting by either mechanism in different patients and causing confusion in the individual case; examples include formaldehyde, isocyanates, and trimellitic anhydride.

Once irritant-induced new onset occurs or aggravation of preexisting asthma becomes severe, the clinical picture is no longer specific. Any low-grade irritant then makes symptoms worse, by provoking cough and chest discomfort. Such patients cannot tolerate side-stream cigarette smoke, fragrances, candles, or other sources of irritating exposures, but this is entirely nonspecific. It is a reflection of the underlying airway reactivity, not a specific or generalized chemical “sensitivity.”

Further on the continuum of irritant-induced occupational asthma is reactive airway dysfunction syndrome (RADS). The diagnosis of RADS was originally based on a set of criteria developed by Stuart Brooks, who discovered it. The essential elements are:

- No previous history of respiratory disorder or atopy
- Onset after a single, specific event of exposure

- Exposure to a gas, smoke, dust, or fumes that has irritant potential at a high concentration
- Onset of symptoms within twenty-four hours and persistence for more than three months
- Symptoms consistent with asthma (cough, dyspnea, wheezing) requiring medical attention
- Airflow obstruction on spirometry

(Additional criteria reflected the use of this formulation for selecting patients with the disorder for clinical research.) These criteria are probably overly specific for the disease and therefore not sensitive enough (see Chapter 5 for an explanation of these terms). In other words, they probably exclude some injured workers who probably merit the diagnosis in clinical terms, but those they include are very likely to be true cases of RADS. For example, there is no obvious reason why a history of respiratory disorder or atopy would preclude the diagnosis of RADS in an individual, because there is no reason that someone with preexisting asthma or allergies would not get RADS; such conditions would not protect against RADS. Brooks's criteria make the exclusion only in order to isolate the disorder and to assemble a relatively homogeneous group of subjects for research purposes. This is an example of a case definition developed for research purposes being misapplied as clinical diagnostic criteria. RADS is really a result of airway injury independent of prior history. In a reaction to the restrictiveness of these criteria, some clinicians have swung to the opposite pole and diagnose RADS liberally after any irritant exposure. This is unfortunate because the natural history of RADS is different from other forms of irritant-induced asthma and so the distinction should be maintained.

RADS is properly diagnosed as the sequela, by history, of a single, discrete event sufficient to induce a persistent inflammatory process that manifests itself as the new onset of airway reactivity. The circumstances of exposure are not obscure, subtle, or easily forgotten;

cases often follow confined-space incidents in which exposure is moderately intense.

Although the management of RADS follows conventional treatment for asthma, the prognosis is highly variable. Several years of airway hyperactivity, with asthma symptoms and variable air flow, can be expected, and some patients never recover normal airway function. RADS is often associated with upper airway problems (sometimes called “reactive upper [airway] dysfunction syndrome,” or RUDS), presenting as obstructive sleep apnea.

### Hypersensitivity Pneumonitis

“Hypersensitivity pneumonitis,” known in the United Kingdom as “extrinsic allergic alveolitis,” occurs when a sensitized individual inhales large quantities of an antigen in the form of a respirable dust that is retained in the lung for some period after being inhaled, although it can be induced by small-molecular-weight chemicals that act as haptens (including the familiar isocyanates and trimellitic acid anhydrides). The lung mounts a T-cell mediated immune reaction that persists because the antigen is not rapidly cleared. The characteristic symptoms of hypersensitivity pneumonitis are dyspnea, fever, chills, and cough, developing over several hours or days. Repeated exposure to the same antigen leads to an inflammatory alveolitis, interstitial fibrosis, and ultimately to a restrictive defect. An immediate hypersensitivity reaction may occur together with the delayed reaction in some cases, resulting in wheezing and air flow obstruction in addition to the more chronic or subacute symptoms. Clubbing is common.

There are over 300 known forms of hypersensitivity pneumonitis, many of them quite rare, and undoubtedly more to be discovered in the future. They are usually given highly descriptive names, such as “cheese worker’s lung” (caused by the mold *Penicillium casei*), “chemical worker’s lung” (caused by any one of several chemical agents), and “humidifier/air conditioner lung” (caused by exposure to thermophilic actinomycetes in forced-air ventilation

systems) but are sometimes given more esoteric names, such as “bagassosis” (caused by exposure to molds on crushed sugar cane, which is called bagasse) and “suberosis” (caused by exposure to molds growing beneath the bark of cork trees, inhaled by workers harvesting cork). Common antigens that produce this condition include molds and actinomycetes of many varieties, detergent enzymes, pharmaceutical agents, minute arthropods such as mites, and dust from vegetable matter, such as grain, or animal material, such as aerosolized droppings and proteins in droppings from birds. Confusingly, certain low-molecular-weight, incomplete antigens that also cause occupational asthma may cause hypersensitivity pneumonitis, including toluene and methylene diisocyanate (Figure 17.22) and trimellitic anhydride. Exposures resulting in a hypersensitivity pneumonitis may occur in many settings, but characteristic histories are exposure to dust from the renovation of old buildings, bird breeding, handling any number of powdered products, and performing maintenance work where powdered waste collects.

Farmer’s lung (Figure 17.23) is the best-studied hypersensitivity pneumonitis and was a historical disorder of practical importance in rural areas with damp climates. The exposure involves farmers handling moldy hay. Farmer’s lung is caused by a specific antigen of *Saccharopolyspora rectivirgula* (previously *Micropolyspora faeni*), a thermophilic actinomycete (not a true mold). Tests for serum antibody for farmer’s lung and a few similar hypersensitivity pneumonitides are available from many laboratories as a panel, but there is not a general test for hypersensitivity pneumonitis. A positive test means that the patient is mounting an antibody response to one of the five or so antigens but does not conclusively demonstrate that that particular antigen is causing the pneumonitis, and a negative result only means that the patient is not mounting an antibody response to one of the antigens in the panel. The second most common hypersensitivity pneumonitis is “bird fancier’s lung,” which involves sensitization to bird proteins and is seen most often among pigeon hobbyists.



**Figure 17.22.** Hypersensitivity pneumonitis in a worker sensitized to toluene diisocyanate, which is also a cause of occupational asthma. Reprinted with permission from "Occupational Lung Diseases," February, 1984, *American Family Physician*. Copyright © 1984 American Academy of Family Physicians. All Rights Reserved.

Patients with hypersensitivity pneumonitis characteristically experience acute symptoms of fever, chills, and shortness of breath, then develop a patchy pulmonary infiltrate, reduced diffusing capacity, oxygen desaturation at rest or exercise, and a leukocytosis.



**Figure 17.23.** Farmer's lung, a hypersensitivity pneumonitis. Note reticular pattern and patchy infiltrate. Reprinted with permission from "Occupational Lung Diseases," February, 1984, *American Family Physician*. Copyright © 1984 American Academy of Family Physicians. All Rights Reserved.

These patients usually recover quickly in the hospital or if removed from the environment and deteriorate on returning to the setting as the disease returns, with crackles and persistent shortness of breath. If exposure persists, the pneumonitis becomes chronic and interstitial fibrosis results, leading to honeycombing and a restrictive defect.

This is one of the most rewarding serious pulmonary diseases to identify but often requires detective work. Identification and control of exposure to the offending antigen usually results in complete resolution of a potentially grave disease.

### ***Industrial Bronchitis***

Workers in dusty occupations, particularly steel workers and grain handlers, may develop a nonspecific chronic bronchitis, which in the older literature is called "industrial bronchitis." Cigarette smoking may aggravate the bronchitis, and smoking-induced bronchitis and cough is obviously the leading alternative diagnosis. This used to be thought of as more or less benign, but recent studies have demonstrated reduced lung function and an increased risk of obstructive lung disease among workers exposed to dusts of many kinds, including those previously thought to be of little health significance and known as "nuisance dusts." This disorder has not been as systematically studied as other occupational lung diseases but would be expected to fit the same pattern of higher risk among workers with atopy and a history of allergy as is observed with most airway disorders.

### ***Fume Fevers***

Fume fevers are systemic disorders resulting from transient effects of mediator release after inhalation. There are two common types of fume fever, both involving mixed pulmonary and systemic reactions to inhaled toxic agents. The pulmonary manifestations are transient and consist of fleeting infiltrates. The disease is a self-limited but highly unpleasant reaction, subjectively similar to influenza, developing an hour after exposure to the offending agent and lasting less than forty-eight hours.

Metal fume fever results from exposure to hot metal fumes, known to include zinc, copper, and cadmium and possibly other metal fumes. An acute, influenza-like illness results consisting of nausea, fever and chills, malaise, myalgias, and leukocytosis. Metal fume fever is most often seen when inexperienced welders try to weld or cut metal that is galvanized or of mixed composition. Within the welding trade, many workers drink milk as a folk treatment or prophylaxis against metal fume fever, but there is no evidence for its efficacy.

Polymer fume fever is a similar influenza-like reaction resulting from the pyrolysis products of chlorofluorcarbon polymers, usually when particles settle on cigarettes, burn, and the fumes are inhaled. Polymer fume fever can be prevented by banning cigarette smoking in the workplaces where products containing these polymers are fabricated.

Metal fume fever must not be confused with the much more serious condition of toxic inhalation, which may result from exposure to high concentrations of cadmium or nickel fumes or from high concentrations of volatilized mercury or lead. Polymer fume fever should also not be confused with "meat-wrappers' asthma," a problem of bronchospasm and irritant bronchitis resulting from the inhalation of fumes generated when polyvinyl chloride film wrapping was cut using a hot wire and discussed under irritant occupational asthma.

## **HEARING CONSERVATION**

Loss of hearing has a profound effect. The auditory system is the primary channel of communication by which a person learns speech and monitors the accuracy of their own speech production, develops language skills, and adjusts to the immediate social environment. Inability to hear the sounds of speech progressively isolates an individual from social interaction. This often leads to feelings of paranoia, frustration, anger, intolerance, loneliness, depression, and inadequacy. It may also lead to frustration and impatience on the part of those dealing with the impaired person.

A major cause of hearing impairment in adults is exposure to excessive noise in the workplace. The OEM physician may become involved when testing programs are introduced into the community or when workers are identified as having impairment of hearing. The most constructive role of the OEM physician is to educate the patient on the importance of hearing conservation, to reinforce compliance with personal hearing protection measures, and to support appropriate rehabilitative intervention for the worker.

already suffering from hearing loss that is impairing communication abilities.

The hearing mechanism is a series of energy transducers, converting energy from acoustical to mechanical, to hydromechanical, and ultimately to electrochemical forms of nerve impulses that can be handled by the brain (Table 17.5). The weakest link in the chain is the conversion of hydromechanical energy to neuronal stimulation. The site where noise exposure is most likely to cause damage is the organ of Corti, in the cochlea. Each delicate hair cell and the neuron or neurons it stimulates is "tuned" to respond best to one fixed frequency. Exposure to repetitive and overly intense continuous noise injures and eventually destroys the delicate hair cells of the auditory sensory receptor.

Hearing impairment is classified clinically as conductive, sensory, neural, mixed, or central. (This is an oversimplification because some conditions such as presbycusis may have components of one or more types of loss.) Special audiologic studies are needed to distinguish among sensory, neural, and central auditory problems. Noise-induced hearing loss is sensory, with some component of neural.

Occupational noise-induced hearing loss is a relatively pure sensory impairment and follows a predictable sequence of events. First, there is a "temporary threshold shift" (TTS) in sound detection. Immediately after exposure and for minutes to hours afterward, the exposed subject cannot hear at the usual level of sensitivity. The TTS phenomenon usually affects frequencies between 3000 and 6000 Hz, which tend to be those about half an octave above the primary frequencies found in most environmental noise. A broader spectrum may be affected with exposure to more intense noise. When the subject's hearing acuity no longer recovers, it becomes a "permanent threshold shift" (PTS). In industrial situations, PTS centers around 4000 Hz. The loss of hearing acuity becomes greater with continued exposure and affects a broader range of frequencies, particularly on the low end of the spectrum. The impairment may become so severe that the individual cannot hear adequately in the critical range of

**Table 17.5.** Conversion of Energy in the Hearing Mechanism

Anatomic Location	Form of Energy	Medium of Energy Conduction	Type of Impairment If Interrupted
<i>Outer ear</i>			
External auditory canal	Acoustic pressure waves	Air	Conductive
Tympanic membrane	Mechanical vibration	Diaphragmatic motion	Conductive
<i>Middle ear</i>			
Ossicular chain*	Mechanical linkage	Mechanical	Conductive
Oval window	Mechanical displacement	Mechanical to fluid	Conductive
<i>Inner ear</i>			
Cochlea	Hydraulic wave	Cochlear fluid	Sensory
Organ of Corti	Nerve receptor stimulation	Neuronal transmission	Sensory
Acoustic nerve (cranial nerve VIII)	Nerve conduction	Neuronal transmission	Neural
<i>Central nervous system</i>			
Brainstem, temporal lobe			
Auditory cortex and association cortex	Nerve conduction	Neural and synaptic	Neural

\* The malleus, incus, and stapes.

frequencies needed to interpret human speech, 500–2000 Hz. At this point, the individual is severely disabled and may no longer be employable. The rate at which noise-induced hearing loss worsens is determined by the level of noise exposure, the duration of exposure, and individual susceptibility, which cannot be predicted.

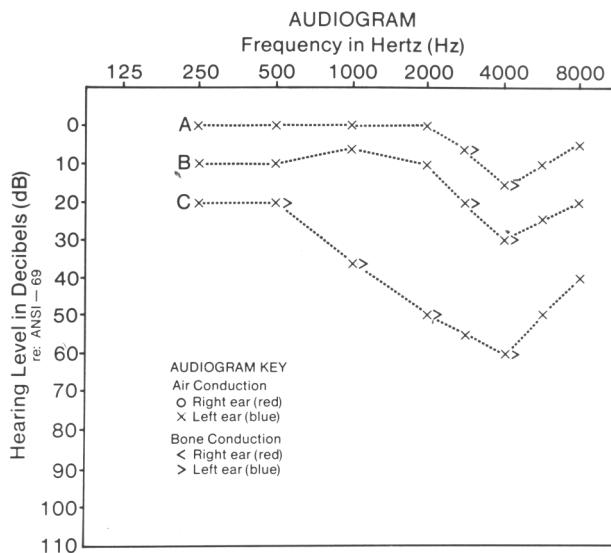
Noise that may be damagingly loud can be recognized by several characteristics: (1) normal conversation levels of speech cannot be maintained at a distance of three feet between speakers, (2) tinnitus may occur upon termination of the noise, and (3) a sensation of “dampened” or less acute hearing follows termination of the noise (the TTS).

Hearing is evaluated by determination of hearing threshold levels at octave intervals from 250 to 8000 Hz, resulting in a graph called the “pure tone audiogram.” A pure tone air/bone conduction audiogram can distinguish between conductive and sensorineural loss but cannot positively identify the cause of the impairment. Pure tone audiometric patterns of noise-induced hearing loss are generally characteristic, however, because they are centered on one frequency, 4000 Hz, where the organ of Corti is most susceptible to injury. The profile of noise thresholds is also usually symmetrical for both ears. Unilateral hearing loss is unlikely to be occupational in origin, unless firearms are used on the job or the worker habitually works with a noise source on one side only.

Noise-induced hearing loss usually starts with a distinct sensorineural “notch” in the audiogram centered at 4000 Hz. Over time, this hearing defect deepens and widens with continued exposure. Noise-induced hearing loss is not flat across all frequencies. Figure 17.24 shows examples of pure tone audiograms resulting from noise-induced sensory hearing loss. Audiograms should be performed only after the subject has been off work and isolated from exposure to loud noise (80 dBA or above) at least overnight, to avoid confusion introduced by a TTS.

Audiograms are highly sensitive tests that should be performed only by professional audiologists or certified technicians in acoustic booths or quiet rooms. The certifying agency for hearing conservation specialists is the Council for Accreditation in Occupational Hearing Conservation. Like all specialized laboratory tests, abnormal audiograms should be interpreted by qualified professionals.

In any individual case, hearing loss may be the result of other causes, most commonly presbycusis, otosclerosis, cerumen impaction,



**Figure 17.24.** Typical audiometric configurations for noise-induced hearing loss as a function of years of exposure to high workplace noise levels: A (1-2 years), B (5-9 years), C (40 years). Results are given for one (left) ear only to show progression of the disorder. Reprinted with permission from "Hearing Conservation and Occupation Exposure to Noise," October, 1983, *American Family Physician*. Copyright © 1983 American Academy of Family Physicians. All Rights Reserved.

ototoxic drugs, middle ear effusions, and Meniere's disease. A characteristic audiogram with or without tinnitus in the absence of other symptoms, particularly dizziness, in a worker exposed to noise on the job for several years is rarely anything other than noise-induced hearing loss.

Impact noise produces barotrauma and acoustical injury based on the pressure wave. This may be associated with hearing loss across a much wider range of frequencies.

## CANCER

The common feature of all occupational cancers is that they can be prevented. Because cancer is the second leading cause of years of life lost in developed countries (because on average it affects people

at a younger age than heart disease) and because the risk increases with aging, prevention of cancer has become a priority in modern society.

### **Environmental Causes of Cancer**

Cancers associated with personal choices of lifestyle, such as cigarette smoking, contribute the greatest fraction of preventable malignancies in North America. Cancers associated with occupational exposures, such as chemicals known to be carcinogenic, are a much smaller fraction of preventable cancers but are of particular concern to society since they are all theoretically preventable. Occupational causes of cancer typically require a latency period of over twenty years (there are many exceptions) until clinical appearance of the malignancy. Occupational causes of cancer are preventable and affect certain identifiable groups disproportionately. Some carcinogens associated primarily with occupational exposures, particularly asbestos and radon, may contribute importantly to national or regional cancer rates because of the widespread distribution of the hazard, the large number of workers exposed, the potential for wider exposure among the general population, and a synergistic effect with cigarette smoking. Environmental causes of cancer, such as radon (lung cancer), arsenic (skin, lung, and bladder), polycyclic aromatic hydrocarbons (PAHs, see Chapter 10) from combustion and in particulate air pollution, and background ionizing radiation place many more people at much lower levels of risk.

Passive cigarette smoke leads as the environmental carcinogen contributing most to the cancer burden. Radon and its daughters together are the most important environmental carcinogen of natural origin causing lung cancer. They are considered to be the two most important environmental causes of lung cancer in nonsmokers. (See Chapter 10 for a complete discussion.) Particulate air pollution is also associated with a risk of lung cancer in urban areas, but it is low in magnitude compared to occupational risks and far below

active cigarette smoking. Arsenic, which is normally encountered as a natural contaminant in groundwater, is associated with cancers at multiple sites, as previously noted.

### **Occupational Cancer**

The burden of disease represented by occupational cancer was initially estimated, based on data from the 1970s, at 4 percent. Contemporary estimates are higher, because of the greater number of asbestos-related malignancies, methodological limitations in the original studies by Doll and Peto (in 1981), the original estimates being based on a narrow selection of cancers (of six types), conservative assumptions about the contribution of occupational causes to common cancers, and the attribution of interactive (synergistic) effects to smoking as the primary cause in the Doll and Peto studies rather than occupational exposure. Contemporary estimates face their own uncertainties, however, and there is no consensus. The true picture is complicated because occupational exposures such as asbestos often interact with cigarette smoking to increase further the risk of lung cancer. Factors such as heredity, diet, and possibly medication may greatly modify an individual's risk of cancer following exposure.

Certain occupational groups, however, may have an unequivocal and greatly increased risk of specific cancers, as has been shown historically for pipefitters (asbestos), dye workers (aniline dyes), and chemists. Some occupational cancers are rare in the general population and seldom seen without a history of exposure to a particular carcinogen, such as hepatic angiosarcoma (vinyl chloride) and mesothelioma (asbestos). Most occupational cancers are common cancers in which occupational exposure to carcinogens contributes additional risk against a background rate, such as lung and bladder cancer. Others demonstrate an elevated risk against a low background rate and can be particularly difficult to evaluate, such as individual non-Hodgkin lymphomas and soft-tissue sarcomas, which probably have different occupational associations.

Workers' compensation claims for cancer are underrepresented (based on expected cancers for known causes such as asbestos), usually disputed (except for mesothelioma), and often insurmountably difficult to prove. As a consequence, the true burden of occupational cancer is greatly underestimated. There are many obstacles to identifying occupational cancers in population studies, as noted in Chapters 2 and 3. Cancer risk is highly age specific and rises exponentially after fifty years of age. Latency may be twenty years or more for a solid tumor (less for leukemias and bladder cancer), and the elapsed time overlaps the increase in risk due to age and is further confounded because the effect of a carcinogen also depends on age at first exposure. Misclassification bias, with respect to the cause of death, the diagnosis of cancer, occupational history, and death certificate data are often compounded by inaccuracies due to the age of the worker at the time of diagnosis, since many cases occur after retirement. Exposure assessment on a scale of decades is very difficult, and interactive effects are extremely difficult to quantify, impossible in small populations. Confounding due to smoking, lifestyle, and other occupational exposures is usual. These obstacles to evaluation in populations spill over to the assessment of causation in the individual. A compatible latency period, documentation of exposure, absence of other plausible causes outside of occupation, and consideration of confounding apply to the assessment of individual cases as well. The mere presence of a known carcinogen in the workplace proves little. No one is harmed if a carcinogen is tightly contained or kept from direct contact with the worker.

Table 17.6 lists occupational exposures known or strongly suspected to cause cancer in humans. Many chemicals, such as 4-aminodiphenyl and  $\beta$ -naphthylamine, were known early to be carcinogenic and have been withdrawn from commercial use. Some chemicals, such as dioxin, are known carcinogens in animal studies but do not show the same level of carcinogenic activity in humans, presumably because of differences in affinity for the molecular receptor. Arsenic is known to be a human carcinogen, but animal studies have not reproduced the carcinogenic effect

**Table 17.6.** Known or Highly Suspect Carcinogens in the Workplace

Substance	Where Encountered
Asbestos	See Chapter 10
Silica	See Chapter 10
Coke oven emissions	Steel mills, coke ovens (includes benzene and BCME)
3,3'-Dichlorobenzidine	Pigment manufacturing, polyurethane production
Radium, radon, radon daughters	Subsurface mining, accumulation in houses
4,4'-Methylene-bis(2-chloroaniline) Uranium and radon	Plastics manufacturing; elastomer, epoxy resins, polyurethane foam
$\beta$ -Naphthylamine	Underground mining
Ultraviolet light	Chemical, dyestuffs, rubber industries
Auramine and magenta	Ubiquitous; working outdoors.
Carbon tetrachloride	Dye manufacturing (withdrawn from commerce)
Benzidine	Very widespread
$\beta$ -Propriolactone	Clinical pathology laboratories; chemical dyestuffs, plastics, rubber, wood products
Vinyl chloride	Plastics, chemical, pharmaceutical industries
Chloromethyl methyl ether (CMME)	Petrochemical, plastics, rubber industries
Bis(chloromethyl) ether (BCME)	Chemical industry
Ethyleneimine	Chemical industry, nuclear reactor fuel processing
N-Nitrosodimethylamine	Chemical, paper, textile industries
Chloroprene	Chemical, rubber, solvent, pesticide industries
Trichlorethylene	Synthetic rubber industry
	Solvent and degreasing agent (withdrawn)

(Continued)

**Table 17.6.** (*Continued*)

Substance	Where Encountered
Benzene	Solvent and chemical constituent
Polychlorinated biphenyls	See Chapter 10
Chloroform	Chemical, pharmaceutical, textile, solvent industries
Acrylonitrile	Plastic, textile industries
Leather dust	Leather goods industry
Wood dust	Hard wood furniture industries
Chromate (hexavalent)	Electroplating, metal products, photography
Nickel subsulfide	Mining and smelting
Ionizing radiation	Medical and industrial x-ray
Arsenic	Mining, smelting; groundwater (natural contaminant)
Cutting oils	Machining, metal-working trades
Hydrazine	Mechanical applications, pharmaceutical industry
Ethylene dibromide	Foodstuffs (fumigation), gasoline additive

because of differences in metabolic pathways. At any given time, two or three dozen chemicals are under active review for carcinogenicity, but many more are being evaluated by commercial, industry, or research laboratories.

The PAHs are a particularly important class of chemical carcinogens (see Chapter 10). They are products of low-temperature combustion and found in coal tar and other complex mixed hydrocarbons. They are found in a very wide range of occupational settings, including foundry work, firefighting, steelmaking (historically in coke ovens, which have largely been replaced in steelmaking), and aluminum reduction (as a constituent of the electrodes in “pot rooms”). The risk that they confer for initiation is compounded by their effects as a promoter and by their effect of photosensitizing the skin and enhancing the effect of exposure to ultraviolet radiation,

particularly in roofing and road building. PAHs are also important constituents of tobacco smoke and air pollution, both of which are important in the risk of cancer. They are associated with a very wide range of cancers, including lung, bladder, skin, and kidney. The PAHs demonstrate the drawbacks of taking the Bradford Hill criteria (see Chapter 3) too literally.

The International Agency for Research on Cancer (IARC) is the scientific body that reviews the evidence for carcinogenicity of a compound extensively used by cancer investigators and regulatory agencies worldwide. (See Chapter 2 for a more complete description.) Group 1 carcinogens are accepted as carcinogenic to humans and are specific as to site. Carcinogens in Group 2A are possibly carcinogenic to humans, and those in Group 2B have insufficient evidence for carcinogenicity in humans but may be carcinogenic in animals. The U.S. Environmental Protection Agency supplements IARC's list and adds many suspected carcinogens for purposes of regulation. The California Department of Health has emerged in recent years as another authoritative body, in that it maintains a list of known and suspected carcinogens that is widely used by other agencies.

Occupational carcinogens vary greatly in their potency. Bis-chloromethyl ether (BCME, used in the nuclear industry in preparing ion-exchange columns to refine fuel-grade uranium) is a highly potent carcinogen that induced oat cell carcinoma of the lung in a high percentage of workers exposed to the chemical. The dust of hard woods, on the other hand, is a low-level carcinogenic exposure resulting in carcinoma of the nasal cavity and sinuses, but although this type of cancer is devastating in its effects, the risk is still small in absolute terms. In some industries and occupations, such as rubber tire making and firefighting, respectively, there is an elevation in cancer risk, but the specific carcinogen is not known and may be multiple. In some settings, such as shift work (see Chapter 13), the induction of cancer is thought to be indirect and not based on the mechanisms of chemical carcinogenesis described in Chapter 2.

The “classic” occupational cancers are those reflecting exposure pathways or susceptible organs and include:

- Bronchogenic (lung) carcinoma
- Skin (squamous, basal cell, melanoma)

- Genitourinary cancers (chiefly bladder)
- Leukemia (acute myelogenous leukemia)
- Mesothelioma of the pleura

Lung cancer due to occupational exposures must be detected against a relatively high background rate in the general population and confounding with cigarette smoking. The minimum plausible latency is usually considered to be between fifteen and twenty years, but the latency period is shortened with more intense exposure and in studies of BCME workers averaged about ten years, which is probably close to a biological limit. Lung cancer is associated with many occupational carcinogens, including asbestos, silica, PAHs, vinyl chloride, arsenic, nickel subsulfide, hexavalent chromium, beryllium, radon daughters and other  $\alpha$ -emitting radionuclides, and sulfuric acid mist (presumably from chronic irritation).

Smoking is the major confounding cause of lung cancer and presents another complication in that cigarette smoking has an interactive (synergistic) effect with exposure to many lung carcinogens, including asbestos, silica, BCME, and radon daughters and probably others carcinogens as long as they do not duplicate the dominant mechanism of cigarette smoking. (Polycyclic aromatic hydrocarbons, for example, are part of and probably contribute heavily to the carcinogenicity of cigarette smoke; the combined effect in that case would probably be simply additive.) The interactive effect is itself exposure-related and would act most strongly when both cigarette smoking and occupational exposures were high.

Occupational causes of cancer may cause any common histological type of lung cancer (epithelial cell, adenocarcinoma, small cell, large cell) and at any degree of differentiation and so, with one minor and one partial exception, histology cannot be used to evaluate potential causes. The minor exception, minor because it is unlikely to be seen today, is that BCME is almost exclusively found with oat cell carcinoma. The partial exception is adenocarcinoma, which although elevated following exposure to asbestos and other occupational carcinogens, is the

characteristic cell type of lung cancer when it appears in nonsmokers (which occurs at about a tenth or less of the rate among smokers). It would be difficult to argue that lung cancer in a nonsmoker was spontaneous and unrelated to exposure if it were of a different cell type. However, the risk of adenocarcinoma of the lung is certainly elevated by occupational carcinogens such as asbestos, so one cannot argue that because a cancer is an adenocarcinoma it is unlikely to be occupational in origin.

Skin cancer arising from occupational exposures may involve direct contact with a carcinogen (for example, PAHs on skin), effects on the skin from permeation by circulating carcinogens (such as arsenic and inhaled PAHs), and photosensitization to ultraviolet radiation (a factor in intense exposure to PAHs). Occupations involving exposure to PAHs, such as roofing, combine all these elements of risk. Occupational skin cancer due to chemical exposure or ultraviolet radiation is squamous cell or basal cell, which is called Bowen's disease if it is *in situ* and associated with arsenic. Mineral oils and ionizing radiation are also known to induce skin cancer. Melanoma is associated with ultraviolet radiation and is unlikely to be caused by chemical carcinogens.

Bladder cancer is associated with occupational exposure because filterable carcinogens concentrate in the urine and expose the bladder epithelium for prolonged periods. The transitional epithelium of the bladder appears to be somewhat prone to initiation, and malignant tumors of the bladder have an unusually short latency period for a solid tumor. Latency as short as seven years was reported among aniline dye manufacturers in the early twentieth century under conditions of very high exposure. Bladder cancer has been associated with a number of chemical dyes that have since been withdrawn, except for benzidine, and with polycyclic aromatic hydrocarbons and arsenic. Bladder cancer is known to be elevated among workers in the rubber industry and among firefighters.

Leukemia is induced by ionizing radiation and alkylating agents. Exposure to benzene is associated acute myelogenous leukemia, and the weight of evidence suggests that this association is highly specific.

Mesothelioma is described below as an asbestos-associated cancer.

Cancers at other sites may be associated with occupational exposure as well and a comprehensive list is outside the scope of this book.

### **Asbestos and Cancer**

Asbestos is historically and quantitatively the single most important occupational carcinogen. The two characteristic cancers associated with asbestos are bronchogenic carcinoma of the lung and pleural mesothelioma. There is strong evidence that other cancers are associated with asbestos exposure, however, including cancer of the larynx, peritoneal mesothelioma, and, less strongly, colon cancer. Ingestion of expectorated fibers, lymphatic drainage, and macrophage migration to the peritoneum are thought to be plausible delivery mechanisms bringing the fibers to other target organs.

### ***Lung Cancer***

Individuals with asbestosis are at high risk of developing lung cancer, but persons exposed to asbestos may certainly develop lung cancer without having asbestosis. Radiological findings of asbestosis are not necessary to conclude that the risk is elevated in an injured worker or claimant. There is some evidence that chrysotile may be somewhat less potent than amphiboles in inducing lung cancer, mostly because fibers may be dissolved over time in the lung, but there is no doubt that chrysotile induces lung cancer. Asbestos exposure is highly interactive (synergistic) with cigarette smoking in causing bronchogenic carcinoma. (Some investigators disagree with each of the points in this paragraph, but the evidence does not support their position.)

The risk of lung cancer is associated with asbestos exposure, and a higher risk is associated with prolonged exposure. Lung cancer requires much less exposure to asbestos than does asbestosis but more than mesothelioma. The latency period between exposure and

detection of the malignancy is on the order of two or three decades. In evaluating a possible occupational association with a patient's malignancy, a complete and accurate history of the patient's employment and the specific jobs performed is critical. Asbestos and smoking together raise an individuals' risk of cancer much more than the sum of either alone, to as much as 100 times that for nonsmokers in heavily exposed, smoking populations. The interaction is proportional to exposure to each agent. Current cases of lung cancer in asbestos-exposed workers reflect distant exposure at much higher levels and also the higher proportion of smokers and consumption of cigarettes in the past. However, current exposure to asbestos in construction and asbestos tear-out work that is noncompliant with occupational standards, a common situation, is high enough to place contemporary workers at risk for lung cancer.

Lung cancer associated with asbestos is not different in its clinical features from lung cancer from smoking cigarettes. It does carry a somewhat worse prognosis than lung cancer in the general population (which of course is usually caused by smoking) and is generally fatal within five years. Cancers in the lung, with a few exceptions, grow for long periods undetected before they become large enough to be seen on x-ray or grow into an airway and shed cells that can be detected. (Figure 17.25 illustrates an advanced lung cancer against a background of asbestosis.) There have been many trials to evaluate the potential benefits of screening for early detection and treatment. So far, they have succeeded in identifying the cancer earlier, but none has successfully demonstrated a benefit in reducing mortality. The net effect is that the injured worker simply knows about the cancer earlier. At the time of this writing, the findings of a major trial performed in Canada using helical HRCT were about to be released.

### ***Mesothelioma***

Pleural mesothelioma is a particularly aggressive and presently incurable malignancy that is closely associated with asbestos exposure. The five-year survival is essentially nil, despite numerous

clinical trials. There are promising new avenues for early detection of this sarcoma through screening, such as circulating levels of a phosphoprotein marker called osteopontin, which may make



**Figure 17.25.** Lung cancer in a sixty-five-year-old smoking shipyard worker exposed to asbestos while installing lagging (insulation). Presented with recurrent laryngeal nerve palsy and hoarseness; lesion was a poorly differentiated squamous cell carcinoma. (Reproduced with permission of the American Academy of Family Practice.)

earlier intervention practical. At present, however, the prognosis is dismal.

Mesothelioma is specific to exposure to asbestos and similar asbestos-form fibrous silicates (such as fibrous zeolites). The mechanism is probably highly nonspecific because it can be induced experimentally in animals by implanting almost any fiber in the pleural space, including cotton threads. Asbestos fibers are carried to the pleural space by lymphatic channels. Although there appears to be a familial predisposition to the cancer, it has been controversial whether it ever occurs in the absence of exposure to asbestos or other fibrous silicates.

Pleural mesothelioma is rare in the general population, but is frequently found in populations exposed to high levels of asbestos. Current cases in developed countries mostly reflect exposure in the distant past, when levels were much higher, but because of the often long latency of mesothelioma, developed countries are currently at or near a peak incidence of the disease. Developing countries, on the other hand, are seeing dramatic increases in rates of mesothelioma due to more recent imports and use of asbestos and will see the peak of their epidemic in coming years. (This has sometimes been called the “second wave” of asbestos-related cancer.) Peritoneal mesothelioma, a rarer but equally insidious disease that declares itself late and may progressively encase abdominal structures in a fibrous tumor mass, likely results from migration of asbestos fibers to the peritoneal space.

Mesothelioma often results from remarkably modest and short periods of exposure. Although it normally has an unusually long latency period of several decades, the latency is quite variable and can be short in the case of early exposure. It is not associated with smoking.

Pleural mesothelioma usually presents as a pleural effusion, characteristically but not always bloody, and a thickened, irregular pleural peel that invades the lung and chest wall (Figure 17.26). The tissue diagnosis is often difficult because these tumors have a highly variable morphology and large blocks of tissue may be needed to differentiate mesothelioma from a poorly differentiated adenocarcinoma. There are different tissue types of mesothelioma, but identification of them is not clinically useful.

The pleural peel is not in any way related to pleural plaques. Mesothelioma should not be confused with either pleural plaques, which are only markers of past exposure to asbestos and signify no additional risk for malignancy in themselves, or benign pleural effusions associated with asbestos.

Mesotheliomas occur spontaneously in the general population, but they are quite rare and many investigators believe that all cases reflect some past exposure to asbestos. Community residents have developed



**Figure 17.26.** Pleural mesothelioma in a sixty-six-year-old nonsmoking retired shipyard maintenance supervisor. After thirty-seven years of employment, during which his exposure to asbestos was not intensive, he retired; two years later he became short of breath while carrying a suitcase and was found to have a massive pleural effusion.

the disease after playing on asbestos tailings as children. There have been clusters of mesotheliomas in families living near sites where asbestos was used or in which a family member was employed in an occupation involving exposure to asbestos and wore his work clothes home, an example of passive exposure.

The amphibole asbestos fibers appear to be most potent in inducing mesotheliomas, particularly crocidolite. The evidence suggests that chrysotile is weak in this regard but certainly not inactive.

## **REPRODUCTION AND DEVELOPMENT**

Reproductive health effects interfere with the ability to conceive and carry a fetus to term. Developmental health effects result in a malformed fetus and present as birth defects. Issues involving reproduction and development are highly visible and often controversial because they evoke emotional responses. They often appear to be more complicated than they are. Most reproductive and developmental hazards have been discovered in animal studies.

The principles of reproductive and developmental toxicology are outlined in Chapter 2. Both generally occur as a result of moderate levels of toxicity; highly toxic exposures would be lethal to the mother and fetus. Reproductive health effects occur against a background of normal fetal loss and the reality that probably less than half of all conceptions continue to a normal delivery. Birth defects occur against a background of 2–3 percent, much higher than most people believe. In general terms, the male reproductive tract is much more susceptible to toxicity than the female. Exposure to an active teratogen (agent that causes birth defects) during gestation must occur at a certain time in organogenesis to cause developmental defects.

Most concern over reproductive hazards arises from the need for protection rather than from clinical problems that require an investigation into cause. Relatively few patients present for evaluation of infertility suspected to be due to occupational or environmental exposures. Relatively few occupational exposures have been found in

practice to cause reproductive toxicity (which usually results in infertility) or developmental toxicity, although certain historical examples (the long-withdrawn pesticide dibromochloropropane, DBCP, and the drug thalidomide, for example) are much discussed models. The principal occupational management problem for reproductive hazards involves risk during pregnancy. The principal environmental issue is the emerging risk of endocrine-disrupting agents, mostly in consumer products and possibly water (which is clearly the major medium for ecological effects). The principal issue with respect to developmental risk at present is associated with pharmaceuticals, not occupational or environmental exposures, although research continues.

OEM physicians frequently find themselves managing issues involving worker protection from reproductive hazards, especially during a worker's pregnancy and often in response to a concerned supervisor. In contemporary North America, these concerns are particularly associated with healthcare institutions, where there is a history of inadequately controlled reproductive hazards (specifically cytotoxic drugs, ethylene oxide, and waste anesthetic gases) and where the workforce is mostly women. The management of women workers during pregnancy is a highly sensitive issue, but the basic rule is that the workplace should be safe enough for them as for all workers. If a job is not safe enough for a woman in pregnancy, it is an unsafe job for anybody and requires attention to hazard control. If there is an unsafe working condition that is an intrinsic part of the job, then guidelines for exclusion or accommodation should apply to all workers (of either sex and with a range of possibility temporary or permanent disabilities) and not solely to women during pregnancy. If the question is whether a pregnant worker can do the usual job, it is simply a question of fitness for duty: matching the job requirements to the worker's capacity to do the job and identifying a reasonable accommodation if there is a remediable gap. Pregnancy is not a disease or an injury, of course, but a normal condition associated with temporary impairment. The extent to which that impairment imposes a disability interfering with performance on the job or safety on the job depends on identifying and monitoring the condition with an

emphasis on capacity—what the pregnant worker can do and finding an accommodation for what the worker cannot do.

The guiding principle, enshrined in U.S. law through the precedent of cases such as *Johnson Controls*, is that enforcing different standards for pregnant workers is discriminatory. Employers are obligated to provide a safe workplace for all workers without regard to sex, pregnancy, or anticipated pregnancy, and occupational hazards should be regulated to ensure safety for all. It is not acceptable to have different sets of standards and to treat pregnant women differently. It is of course prudent to ensure that during pregnancy there is no risk of an inadvertent or uncontrolled exposure.

The interest of workers is that there be no loss of income due to reassignment to a less well-paying job and at the same time that there be no risk to the worker.

As a practical matter, the most common exposures of reproductive concern in the workplace are associated with healthcare and have historically been a risk to women health workers such as nurses, central sterile supply workers, and radiology technicians. These hazards include anesthetic gases, cytotoxic drugs, cytomegalovirus infection, rubella, ethylene oxide, and ionizing radiation. Improvements in hospital and healthcare protection in the 1980s controlled these exposures after a series of alarming revelations regarding the risk of miscarriage among nurses and central sterile supply technicians. Beyond healthcare settings, a slowly increasing number of workplace exposures are suspected of having adverse effects on male or female fertility, birth outcome, fetal development, and congenital abnormalities. Particular occupations, including firefighters and welders, appear to have elevated rates of infertility. A somewhat smaller number show a suggestion of increased miscarriages, historically including oncology nurses. Documented outbreaks of birth defects due to occupational and environmental teratogens are very rare. Attributing an adverse reproductive outcome to a particular exposure is further complicated by exposure in daily life to potentially fetotoxic agents, such as some therapeutic drugs, cigarette smoke, and alcohol.

Many suspected reproductive and developmental hazards have been ruled out. Contrary to allegations in the 1980s, there is no risk from exposure to video display terminals, for example.

Physical hazards, specifically heat, vibration, pressure, and ionizing radiation, probably represent a greater management problem in practice than chemical hazards. Hyperthermia is potentially teratogenic and so excessive heat should be avoided during pregnancy. Total body vibration is poorly studied but may represent a risk factor for premature delivery and should be avoided. Decompression illness after diving, following work in hyperbaric atmospheric pressure followed by depressurization, represent a risk to pregnant women and especially the fetus due to gas embolism and diving should not be undertaken during pregnancy. Ionizing radiation is, of course, a major hazard to the fetus.

Ergonomic issues can generally be managed with easy accommodations during pregnancy. Moderate physical exertion is not any more of an issue during pregnancy than before it, but heavy exertion should probably be curtailed in the last few weeks before delivery. Heavy lifting has been studied and found not to be a risk factor for adverse birth outcomes and so would not be contraindicated if the woman is accustomed to it and feels able to do the work. Many women find prolonged standing to be a problem during pregnancy, particularly if they are experiencing bloating, retaining fluid, or have developed varices. However, pregnancies are different and the management of accommodations during pregnancy should be worked out between the woman and her employer, not by a rigid standard. Employers may offer reassignment or pregnancy leave, but the worker, once informed, should make the decision about tolerable work conditions.

## SYSTEMIC TOXICITY

The principles of clinical toxicology are outlined in Chapter 2. The toxicity of several common agents is presented in Chapter 10. This section highlights a few selected but important situations

involving systemic toxicity. Confined-space incidents and “man down” situations are usually discussed in terms of safety and prevention, which in the workplace falls into the domain of the safety officer. However, an understanding of these situations is most important for the OEM physician when he or she encounters a patient with such a history or is called in an emergency.

### **Confined-Space Incidents**

Confined-space incidents occur in enclosed areas not usually occupied by workers where gas accumulates or oxygen levels diminish, creating a hazardous atmosphere. There may also be mechanical hazards, electrical hazards, and obstacles to entry and exit. By definition, confined spaces are hard to ventilate, and workers caught in them may find it difficult to escape and especially to turn around in a narrow crawl space or to climb out on a ladder if access is overhead. Likewise, access may be difficult for rescuers. Common confined spaces include empty tanks, utility rooms, holds on board ships, wells, cisterns, refinery vessels, reaction chambers, and pits.

Within a confined space, toxic gases such as hydrogen sulfide may accumulate to high concentration without ventilation. Flammable or explosive gases may accumulate, for example, from solvents used to clean the inside of a tank. Relatively inert gases may be used to purge vessels (especially if they have held hydrocarbons or other potentially explosive gases), displacing oxygen. Welding inside the confined space may produce oxides of nitrogen and deplete oxygen levels. Oxidation or other chemical reactions may consume and deplete oxygen in the confined space. Gases and vapors that are heavier than air may accumulate in tanks that are open at the top, below decks in ships, in storage chambers, and in subsurface depressions, such as pits, holes, and trenches. Carbon monoxide, hydrogen sulfide, and hydrocarbon vapors may collect in such spaces, with lethal results, when emission sources are near the entrance to a confined space and can funnel in by gravity or air movement. A particular problem in agricultural and rural areas has been hydrogen sulfide toxicity in

liquid manure collection pits and septic tanks. Heat and cold stress are potential hazards in some situations.

A typical scenario has a worker entering a confined space due to ignorance, impatience, or a false report of safety and then collapsing. A second worker, on observing or discovering the collapse of the first usually attempts a rescue unprotected by breathing apparatus, only to succumb. These incidents are often fatal and almost invariably come in twos (and sometimes threes or more) because of ill-advised heroic rescue attempts. They are a major cause of “man down” events (as described below).

Any setting in which a confined-space entry may occur requires that the workers be trained, that proper protective equipment be available, and that procedures are in place to prevent unauthorized entry and solo work in confined spaces. OSHA has confined-space standards for general industry, construction, and shipyards that must be followed in the United States.

Employers in industries in which this is a problem should have confined-space entry policies requiring certification of workers who are trained to work safely in confined spaces and who are issued a permit to enter the specific space. The confined space should be tagged and locked down at all other times. Entry should occur only after the atmosphere has been tested for oxygen level, combustible gases, and finally toxic gases that may plausibly be present, in that order (because direct-reading instruments for combustible gases require oxygen in the atmosphere for reliable readings). An oxygen level of 20 percent is required for safe entry; lower levels, whether or not they are high enough to sustain life, are associated with impaired judgment, impaired hand-eye coordination, and strength, making work unacceptably dangerous and impairing the worker's ability to escape.

Confined spaces should always be ventilated when work is being done, providing both air supply and exhaust channels. Supplied-air respirators should be used by workers entering the confined space if there is any question about the safety of the atmosphere and whether it might change (for example, if there is a leak of gas or a connection with a sewer that could result in entry of sewer gas). Air purifying

respirators should never be used in confined spaces where the atmosphere presents a potential hazard of oxygen depletion or where the hazard could possibly exceed the protection factor of the respirator. Self-contained breathing apparatus, ropes, and harnesses should be immediately at hand near the confined space at all times for use in rescue efforts. Only trained personnel should be allowed to enter confined spaces, always using the “buddy system” of working with a competent companion capable of alarm and rescue actions who stays outside and is in constant contact with the worker inside the space.

### **“Man Down”**

In industry, the usual term for a worker who appears to be unconscious and to have suddenly collapsed is “man down,” a historical phrase borrowed from the military and emergency services which should be considered gender neutral. “Man down” situations should always be considered to be life threatening until proven otherwise.

As the term implies, “man down” calls involve a worker (or visitor) who has collapsed and cannot get up.

Rescue is primary, but the rescuer himself (it is almost always a man) must be protected (see Chapter 14) before attempting to remove the unconscious worker to safety from a confined space and, if there is any possibility of continued exposure, must be equipped with a supplied-air respirator. These incidents usually result in multiple casualties because of ill-advised attempts at heroic rescue.

CPR should obviously be initiated immediately if the patient is apneic and pulseless and the usual clinical measures pursued, including an immediate blood glucose determination. If there is a possibility of cyanide exposure, the rescuer and first responders must protect themselves from exposure due to contamination. The OEM physician should consider the usual causes of syncope and the possibility of a toxic or other occupational cause in the context of that specific workplace.

The cause may be obvious, such as a sour gas well blowout (hydrogen sulfide), an air compressor too close to a source of

combustion supplying an air line (carbon monoxide), or anoxia in a confined space (oxygen deficiency), but sometimes it is not. In the absence of trauma or electrocution, the cause of a “man down” is most likely to be a health event, such as an arrhythmia, myocardial infarction, or overwhelming toxic exposure (see Chapter 9). Hypoglycemia in an insulin-dependent diabetic is possible but is most likely to have occurred before, and brittle diabetics usually wear alert bracelets. Vaso-vagal syncope, which might occur in an acutely stressful condition, with prolonged standing in place, or with heat stress, would be unusual in a healthy worker under normal circumstances.

Certain causes of loss of consciousness can usually be ruled out quickly: Seizures are usually obvious. Orthostatic hypotension corrects itself when the head reaches the level of the heart and the patient will normally revive quickly; it is unlikely in the workplace except among nonacclimated workers exposed to heat stress.

The most likely occupational exposures that could be responsible for this relatively uncommon pattern of collapse are asphyxiants (gases that displace oxygen in a confined-space situation as above), tissue asphyxiants (chemicals interfering with oxygen delivery and cytochrome metabolism), cardiotoxic agents, and anesthetic-acting vapors.

True “asphyxiants” are any gases, usually relatively inert, that displace an atmosphere containing oxygen, including hydrocarbons, chlorofluorocarbons, carbon dioxide, and the true inert gases, including nitrogen. This situation almost invariably occurs in the context of a confined-space incident. The problem is not one of toxicity but oxygen deprivation.

Tissue asphyxiants are agents that interfere with oxygen delivery or energy metabolism by inhibition of cytochrome oxidase or hemoglobin. There are four gases of significance as tissue asphyxiants, forming a spectrum as to their degree of toxicity to cytochrome and hemoglobin on the one hand and their degree of irritation to tissue, principally mucosa, and bronchial epithelium on the other. These are, in order, carbon monoxide (potently competing with oxygen for binding to hemoglobin, not at all irritating), cyanide (potent

inhibitor of cytochrome oxidase, irritating), hydrogen sulfide (weakly inhibiting cytochrome oxidase, highly irritating and a cause of sudden loss of consciousness), and azide (weakly inhibiting of cytochrome oxidase, very irritating, and a cause of sudden hypotension). Carbon monoxide, cyanide, and to a lesser degree hydrogen sulfide act in different ways by interfering directly with oxidative metabolism at the cellular level. They are discussed more fully in Chapter 10. Cyanide contamination may place rescuers, first responders, and other healthcare providers at risk during unprotected mouth-to-mouth resuscitation and while handling contaminated vomitus. Hydrogen sulfide has the effect of a sudden, complete loss of consciousness, popularly called a “knockdown,” which is rapidly reversible if exposure ceases, and has other effects associated with eye irritation, apnea, and mucosal irritation.

Azide is an important preservative in biotechnology and biological research and the critical component of the explosive charge used in automobile air bags. Its toxicity appears to be mostly unrelated to its weak inhibition of cytochrome and results from its potent effects on vasodilation. Like many other nitrogenous compounds, such as nitroglycerin, azide causes rapid venous capacitance dilation and a drop in smooth muscle tone in the walls of arterial and venous blood vessels, abruptly reducing both preload and afterload and precipitating acute hypotension and syncope. Once the person collapses and he or she is prone or supine, the pressure gradient between the heart and the brain equalizes and cerebral perfusion resumes; recovery may take place relatively quickly except for a residual headache and lightheadedness. The result is not quite so abrupt and may be preceded by a pounding headache due to vasodilation.

Cardiotoxic agents that can cause abrupt collapse act by inducing arrhythmia, including ventricular tachycardia or fibrillation. These are chiefly the chlorinated hydrocarbons, including chlorofluorocarbons and chlorinated solvents.

Hydrocarbon vapors, especially solvents, may induce narcosis, like anesthetic gases. This effect requires exposure at high concentrations and occurs in confined spaces.

## Arsine Toxicity

Arsine ( $\text{AsH}_3$ ), and its germanium counterpart germane ( $\text{GeH}_4$ ), are gases used mostly in semiconductor manufacturing. They may cause acute massive hemolysis and rhabdomyolysis due to their affinity for oxidized hemoglobin and myoglobin. Respiratory distress syndrome, acute renal failure, and marrow toxicity may occur simultaneously and almost instantaneously. In nonfatal cases, hemoglobinuria, hepatomegaly, and a drastically reduced red cell count occur. If the victim survives, complications are numerous, and treatment may require exchange transfusions, hemodialysis, and intensive supportive care. Blood or urinary arsenic is a useful aid to diagnosis and a means of following progress in arsine poisoning, but the diagnosis must be made immediately and clinically. Arsine toxicity has occurred historically in the semiconductor and smelting industries.

## OCCUPATIONAL NEUROLOGY

Neurological conditions associated with acute toxic exposures are discussed elsewhere, with respect to the causal agent.

Three neurological conditions in occupational medicine are featured in this section: toxic encephalopathies, Parkinsonism disorders, and toxic peripheral neuropathies. Treatment of toxic neuropathies has been limited in the past to facilitating spontaneous recovery by removing from the patient further toxic exposure (and reducing body burden of heavy metals by chelation, if necessary) and nutritional supplementation (specifically using vitamin  $B_6$ , which unfortunately can also induce a neuritis at high levels of consumption).

## Toxic Encephalopathy

High levels of exposure to variety of agents are capable of causing central nervous system toxicity. Some of these effects are discrete, as in the case of manganese. Encephalopathy is a general condition manifested as behavioral disturbance, changes in mood, cognitive

impairment, and diminished short-term memory. Unfortunately, the term “toxic encephalopathy” has been appropriated and trivialized by some advocates associated with the “environmental sensitivity” movement, but it properly means a condition of gross functional impairment in the brain associated with toxic and metabolic causes. Toxic encephalopathy is a medical emergency requiring immediate diagnosis and intervention.

Lead has been the most common cause of toxic encephalopathy, historically, in occupational medicine. Lead encephalopathy in adults occurs at blood lead levels above 100 µg/dl and in children at levels around 80 µg/dl, which are catastrophically high (see Chapter 10). In children, lead encephalopathy is associated with projectile vomiting, irritability, hyperactivity, ataxia, visual changes, seizures, and a peculiar passive state that has been characterized as wakeful lethargy; it progresses to coma and has a very poor prognosis, fatal in 25 percent of the cases, and with residual neurobehavioral impairment in most survivors. Whether cognitive impairment of this severity occurs in adults, short of massive exposure, is controversial, but chronic lead exposure in adults is associated with impairment in attention, short-term memory, and psychomotor skills, as well as irrational behavior, labile effect, irritability, and apathy. Lead encephalopathy is a true medical emergency, whether in the adult or child and is treated by a combination of BAL and CaNa<sub>2</sub>-EDTA (never EDTA alone, because of the risk of inducing fatal hypocalcemia), but the need for this is rare. Because cerebral edema is part of lead encephalopathy, lumbar puncture is contraindicated. The edema is treated with mannitol (and other hyperosmotic treatment), steroids, and hyperventilation if required. Fluid management is necessary because cerebral edema may result in the syndrome of inappropriate ADH, and seizures are treated with benzodiazepines.

Tetraethyl lead, the additive in leaded gasoline, was the most potent exposure historically associated with severe lead encephalopathy and was responsible for an outbreak of fatal and disabling encephalopathic cases at the Ethyl Corporation in the 1920s, which presaged the hazards of leaded gasoline.

Among its other effects, arsenic may also cause a syndrome of malaise, labile affect, and neurobehavioral effects and is also recognized for neurocognitive effects in children at much lower levels.

The encephalopathic effects of mercury are dramatic and fortunately uncommon: extreme emotional lability, personality change (often characterized as pathological shyness), cognitive dysfunction, attention deficit, and irritability. The syndrome of mercury-induced encephalopathy, also called “mercurial erethism,” is familiar in the expression “mad as a hatter” and in the characterization of the hatter character in Lewis Carroll’s *Alice in Wonderland*; hat makers in the seventeenth and eighteenth centuries were at risk of mercury intoxication from contact with mercurials used as antifungal agents on the felt cloth used to make hats. Metallic mercury can be absorbed by the inhalation route, since it is a volatile substance, in quantities sufficient to cause toxicity (Chapter 10). The greatest concern over inhalation of mercury has been in the exposure of dental assistants, research laboratory personnel, and instrument and measurement equipment workers, all of whom may work with mercury under circumstances in which spills and prolonged exposure is possible. Organic mercury compounds are even more toxic; a severe outbreak of methylmercury poisoning from the consumption of food and water contaminated by a chloralkali plant on Minimata Bay, in Japan, received worldwide attention. The methylated, organic derivative of mercury is more easily absorbed by the body and is much more toxic; it is this form that is present in river and lake sediments in which mercury has been discharged into the environment.

Organic derivatives of tin, particularly the marine antifouling paint additives triethyl tin and trimethyl tin, are potent central neurotoxins. In animals and in the few human cases studied, trimethyl tin produced behavioral and affective changes, short-term memory loss, and hearing impairment. The limbic system appears to be particularly affected.

Among organic compounds, carbon disulfide ( $\text{CS}_2$ ) is associated with a severe and characteristic encephalopathic condition, which includes the usual features but also, and sometimes in isolation, an

induced thought disorder manifested by paranoid ideation, manic behavior, and suicide. With the introduction of improved containment and worker protection in the rayon industry and the reduction in use of CS<sub>2</sub> as a fumigant, the syndrome is rare today.

Solvent exposure may produce encephalopathic changes; toluene and trichloroethylene are most often implicated as single agents, but exposure to multiple solvents is usual and the problem is difficult to study. Many investigators believe that effects are not limited to these two specific agents. The specific neurotoxicology of the solvents is discussed in Chapter 10.

Organophosphate pesticides are associated with encephalopathic changes acutely, as described in Chapter 10. Many clinicians have suggested that persistent changes may follow recovery from acute exposure, manifested by labile affect, attention deficit, anxiety, memory deficits, headaches, and irritability, but this has not been confirmed.

## Movement Disorders

Manganese produces a syndrome closely resembling Parkinson's disease, with an extrapyramidal motor disorder associated with injury to basal ganglia and the characteristic mask-like facies. Exposure is by inhalation. Figure 17.27 presents a woman who was exposed while working as an investigator conducting research on manganese compounds at a metallurgical research institute in China and who subsequently developed a florid condition identical to idiopathic Parkinson's disease. The disease is also found among welders who have welded on ferromanganese alloys. Prognosis is poor as the toxicity is irreversible.

Parkinsonism is also associated with intense solvent exposure and possibly pesticide exposure, but the specific solvents and pesticides that may be responsible are not clear. A related neurological condition known as "subacute combined system atrophy" or degeneration (not to be confused with combined degeneration of the cord, which occurs as a result of vitamin B<sub>12</sub> deficiency) is an autonomic disorder characterized by orthostatic hypotension, and dementia may also be associated with occupational exposure to solvents and possibly pesticides.



**Figure 17.27.** Parkinson's disease-like neurological condition in a Chinese metallurgical research scientist heavily exposed to manganese. Ferromanganese alloys are extensively used for rails, and most cases in China are seen in railroad welders.

## Toxic Peripheral Neuropathy

Peripheral neuropathies may be predominantly or purely motor, sensory, or mixed. Motor neuropathies present as palsies, muscle weakness, and atrophy and diminished tendon reflexes. Sensory neuropathies usually present as dysesthesias, often of a burning quality, and numbness to pain and vibration, usually in a "stocking-glove" distribution. Neuropathies affecting autonomic nerves may result in characteristic malfunction of sphincter activity, orthostatic hypotension, and dystrophic changes over the skin. Nerve conduction studies, assessing conduction velocity and amplitude of response, are useful confirmatory tests for neuropathies affecting accessible nerves. Electromyography is less often helpful.

It is difficult to be sure that a peripheral neuropathy is due to an occupational cause in a particular case. The disorder occurs most often in association with diabetes, as a result of alcohol abuse, and idiopathically. Peripheral neuropathies are side effects of common drugs, including isoniazid, penicillin, phenytoin, furadantoin, and metranidazole, and with several antineoplastic drugs, including vincristine, adriamycin, procarbazine, and 5-fluorouracil. Consumption of excessive quantities of vitamin B<sub>6</sub> may cause a neuropathy. There are numerous other less common causes of peripheral neuropathy, among them porphyria, amyloidosis, hereditary conditions, collagen vascular disorders, and as a systemic effect of cancer.

Usually, there is a history of exposure to a known neurotoxic agent over a period of months or years, with gradual onset of the neuropathic symptoms. (An exception is Vacor, a neurotoxic p-nitrophenylurea rodenticide that resembles streptozotocin and may cause symptoms of peripheral neuropathy, diabetes and dysautonomia within days.) In the case of pesticides, exposure to the organophosphate pesticides may precede within two weeks of the onset of a primarily motor neuropathy.

Metals, especially lead, have been the most common cause of toxic peripheral neuropathy in occupational settings historically (see Chapter 10). Lead characteristically causes a "dying-back" motor axonopathy affecting the muscles of the forearm innervated by the

radial nerve. The resulting extensor nerve palsy and unopposed flexor action causes the condition known as “wrist drop,” seriously impairing use of the hands. Mercury is associated with tremor and motor disorders. Thallium also causes motor neuropathy, but toxicity is usually obvious because of patchy hair loss. Bismuth and arsenic may cause sensory neuropathies. Copper, zinc, gold, silver, and tin may cause mixed sensory or motor neuropathies, but these are rare.

The classical historical causes of peripheral neuropathy among organic chemicals have disappeared from the workplace: *n*-hexane and methyl-*n*-butyl ketone, both of which are metabolized to the highly neurotoxic product 2,5-hexanedione (see Chapter 10). The so-called  $\gamma$ -diketones form a pyrrole in the cell that eventually cross-links neural filaments in the axon, resulting in axonal swelling, a “dying-back axonopathy,” myelin swelling, and a progressive neuropathy that continues after exposure ceases, a phenomenon called “coasting.”

Other solvents, including trichloroethylene, benzene, carbon tetrachloride, and carbon disulfide, have been associated with peripheral neuropathy, but these are rare in the contemporary workplace. Exposure to tetrachloroethane (perchloroethane) has also been suggested as a contributing exposure potentiating neurotoxicity from other solvents. 1-bromopropane, a solvent introduced to replace ozone-depleting chlorinated solvents, has been associated with a severe neuropathy characterized by pain and spastic paraparesis.

Organophosphate esters may cause an acute or a delayed neurotoxicity, as described in Chapter 10 for organophosphate pesticides. Tri-ortho-cresyl phosphate was an important neurotoxin historically, responsible for a major outbreak of “ginger jake” motor paralysis during Prohibition, when flavor extracts adulterated with the substance were consumed in large quantity for their alcohol content. Some organophosphate pesticides may cause axonal changes and motor paralysis of the lower extremities as a long-term sequela: leptophos, mipafox, and trichlorphon in particular. The effect appears to result from inhibition of a specific esterase other than the acetylcholinesterase at the post-synaptic junction or neuromuscular junction responsible for the acute effects. This

“neuropathy target esterase” is of unknown function. The organophosphate-induced delayed neuropathy syndrome consists of proximal weakness of limbs and the muscles of respiration, leading to a condition resembling Guillain-Barré syndrome and sometimes requiring assisted ventilation. Onset is usually within one to three weeks of exposure. Another distinct but less well defined delayed neuropathy has been described as having onset twenty-four to ninety-six hours after exposure.

A specifically autonomic syndrome has been associated with dimethylaminopropionitrile (DMAPN), a catalyst formerly used in polyurethane manufacturing. DMAPN affects sacral nerves selectively, causing urinary retention and impotence. The identification of this devastating disorder was a particularly elegant example of astute clinical observation leading to identification of a hazard.

Acrylamide, the monomer of polyacrylamide gel preparations, is a potent neurotoxic agent that also produces a “dying back axonopathy” and induces a combined sensory and autonomic neuropathy with dermatitis and muscle wasting. Some immediate central effects have been reported with acute exposure, including ataxia and delirium. Fortunately, both the chronic and the acute condition appear to be reversible.

## **OCCUPATIONAL INFECTIONS**

Compared to other areas of medicine, OEM deals less often with infectious agents and disease and more often with chemical hazards. However, infectious agents and disease and the risk of exposure to pathogens is an important part of OEM practice and are distinguished from encounters with infectious disease elsewhere in medicine mainly by context, not recognition or medical management.

Infectious agents and disease are part of the training of every qualified physician, and information on them is readily obtainable through the usual medical information sources. The diagnosis and treatment of infectious disease is the same in OEM as in clinical medicine otherwise. Therefore, this section will not go into detail on the individual

infectious diseases encountered in OEM. Rather, it will discuss the context in which they are encountered and the broad management issues of each, apart from diagnosis and treatment.

The contexts in which infectious disease are managed in occupational medicine are:

- Management of pathogens and infectious hazards in the workplace
- Infectious disease arising out of work as a result of biological hazards specific to the workplace
- Infectious disease arising out of work due to location and activity that is not specific to the workplace
- Community-acquired infectious disease and its implications for the workplace
- Accommodation and adjustment issues for workers with chronic disease
- Travel and the risk of acquired infection

The latter two topics are discussed elsewhere in this book.

All of these examples are important for some infectious agents, such as HIV, which serves, unfortunately, as an excellent context for illustrating the issues. HIV/AIDS is a specific risk in the healthcare workplace with specific procedures for primary prevention (for example, preventing needle-stick wounds) and secondary prevention (antiretroviral prophylaxis when there is a high risk that infection may have occurred) in the workplace. It is fortunately uncommon among healthcare workers but critical to identify early when it does occur. HIV/AIDS is related to work in indirect ways as well, as demonstrated by the high prevalence among truck drivers and miners in southern Africa, where the risk reflects transmission patterns due to geographic distribution and lifestyle. It is as or more likely to occur due to exposure outside the workplace, and the risk is proportionate to the prevalence in the relevant general population; once a worker in a sensitive occupation, such as a

healthcare worker, is infected, this information has implications for work capacity, although fewer than for more communicable infections. There may be accommodation issues for individuals with HIV/AIDS undergoing treatment and also work adjustment issues with respect to access to healthcare and maintaining an antiretroviral regimen. Travel to areas with a high prevalence of HIV/AIDS presents an opportunity for increased risk depending on personal behavior, the management of which may involve highly sensitive discussions with the travelling worker (who is often a technical expert, senior manager, or executive and who may often resist intrusive counseling by the physician with respect to sexual behavior). Promotion and even provision of protection (such as condoms) may be required when risk-taking behavior is unlikely to change. Another scenario in which HIV/AIDS may be a travel risk is when medical care is provided under conditions of questionable sterilization or when transfusions are required in high-risk areas. Travelers may need to carry their own sterile supplies, such as needles, to avoid infection during emergency care in high-risk areas. Because vehicular accidents are more common and more likely to result in serious injury in less economically developed countries, as a rule, the risk of HIV infection for some travelers may be driven more by traffic and transportation choices than another factor. Finally, there remain situations in which HIV positivity is a factor in immigration and visa eligibility. HIV/AIDS therefore well illustrates the range of occupational management issues for infectious disease.

Hepatitis B, which is more readily transmissible, reflects the same issues.

### **Infectious Hazards in the Workplace**

In Chapter 2, chemical agents were described as having an exposure-response relationship that, at a certain threshold level or case definition, results in a certain frequency or risk of toxicity. The exposure-response relationship was a gradient, resulting in degrees of toxicity. Other disease types, such as cancer, were described as

“stochastic,” in that they occur with a certain frequency at any given level of exposure to a carcinogen but the disease itself is the same. Infectious agents, which propagate after an initial innoculum, follow stochastic dynamics. This means that the key to controlling infectious disease in the workplace is not so much to reduce the level of exposure, although this is desirable, but to eliminate the opportunity to come into contact with the pathogen altogether. Hospitals and healthcare institutions provide the most convenient example because they are high-risk environments and also familiar to every physician.

Infectious agents are encountered in many settings, but the single most hazardous employment sector overall for infectious disease is healthcare. This is not surprising, since the hospital and clinic bring to one place all manner of patients and medical conditions in close interaction. Whether on the ward or in the waiting room, patients with communicable diseases may encounter each other and healthcare workers, who may or may not carry communicable diseases of their own (such as staphylococcus colonization in their nares), with the potential for transmission. This occurs under conditions dense enough to accelerate transmission in the population as a whole for the more communicable diseases. Healthcare settings are therefore rich environments for disease transmission as well as for the selection of resistant strains by passage through hosts who are being treated. However, healthcare occupations have acquired a false perception of being “clean” because of their environment, and so for many years, until very recently, healthcare lagged as a sector in worker protection. For many years the primary infectious threat in hospitals was hepatitis B, yet even the protean and potentially severe manifestations of that disease was insufficient to motivate adoption of universal precautions. The threats, in sequence, of transmission of HIV, multi-drug-resistant tuberculosis, the SARS agent (not known at the time to be a coronavirus), avian and other potentially pandemic influenzas (“novel influenza A H1N1” being the one circulating at the time of this writing), and methicillin-resistant staphylococcus dramatically changed attitudes in the healthcare sector. They also broadened concern over hazard protection from

an emphasis on hand washing and needlestick injury, to the inhalation route to housekeeping, to rigorous control of all routes of exposure today. OEM physicians with responsibility for healthcare institutions, and especially hospital employee health services, are cautioned to benchmark programs, policies, and performance against both best practices in the healthcare sector and also general industry, where attitudes are different and often more proactive.

Because of the stochastic dynamics of transmission, the objective of infection control in the hospital workplace, used here as an example for workplaces at high risk, is the approach of multiple barriers. Exposure to a chemical hazard is reduced by the protection factor of each control measure (such as isolation, ventilation, engineering controls, personal protective equipment, and administrative or behavioral controls), and risk is reduced proportional to the reduction of exposure and stays reduced as long as the control measures are in effective operation. Infectious agents present a different problem. A lapse in protection against a communicable agent may reduce the probability of transmission, but once protection is breached the agent can propagate or may spread by person-to-person transmission. Isolation of infectious cases, negative air pressure in infectious disease wards and isolation units, universal precautions (which are not described here because they should be intimately familiar to all practicing physicians), antibiotic prophylaxis (for staphylococcal carriers), and meticulous hand washing each provide protection that, although potentially incomplete, reduce the risk in different ways. If a pathogen is not “intercepted” and so blocked by one barrier, it is likely to be blocked by the next. In this context the many modalities for hazard control (which run the gamut of control measures discussed in Chapter 14) are not redundant or duplicative but sequential. Unfortunately many of these control measures depend on individual behavior, making compliance with often complicated procedures the limiting factor for the entire chain of protective measures. A break in any step creates a gap that may result in transmission and spreading, undermining the whole of the infection control effort. The goal is to reduce the cumulative probability of

transmission (a function of the product of individual protection factors of each control measure and their failure rates) to level far below the probability that a single case would ever occur in the protected population.

### **Infectious Hazards in the Workplace**

There are many microbial hazards specific or even unique to certain occupations. The specific workplace hazards are associated with infections in the occupation, such as orf (a spirochetal disease) in sheep handling, leptospirosis among sugar cane workers, anthrax (usually cutaneous) among workers who handle unsterilized cattle hides, and brucellosis among large-animal veterinarians. These hazards are concentrated in occupations that involve animal handling, agriculture, microbiological laboratory services, healthcare, and behaviors that may lead to the exchange of body fluids (obviously rare in the workplace, with the exception of sex workers).

Other than healthcare, most infectious hazards are particular to the occupation and are relatively unlikely to be encountered in a general OEM practice. A somewhat broader range of infections may be encountered among animal handlers with an opportunity to encounter “zoonoses” (diseases carried by animals that can affect human beings). They are more likely to be seen in an infectious disease consultation. Sexually transmitted infections among sex workers are not unique as a hazard to those workers but are not usually managed by OEM physicians and because of social approbation may not come to medical attention early. There are special clinics and healthcare providers for sex workers, but they are not in occupational health centers. For example, there is a clinic in Los Angeles dedicated to the health of performers in the adult film industry; it emphasizes prevention of HIV/AIDS.

Most, but certainly not all, microbial hazards are well known within their industries, and guidance for the OEM physician is usually readily available through the usual sources of medical and occupational health information. For that reason, no attempt will be

made to list the numerous possibilities here. *Couturier's Occupational and Environmental Infectious Diseases* covers the field exhaustively.

### **Work-Related Infectious Disease**

Infectious diseases that are endemic to a particular location or setting within which work is performed constitute a much greater challenge for OEM physicians than unique infectious hazards in the workplace, outside of healthcare. Much of occupational medicine as it is practiced today arose from the need to control tuberculosis in the workplace at a time when it was highly prevalent in the community. The basic principles of pre-employment evaluations (before pre-placement evaluations replaced them) and periodic health surveillance were worked out to screen workers for the disease and to prevent transmission within the workplace, using the limited tools of the day.

Today the single largest health challenge associated with location and workplace setting is control of malaria, the most prevalent infectious disease in the world. Employers with operations that involve outdoor work in malaria-endemic areas, such as heavy construction and oilfield work, face a constant battle with the disease in their workers and the risk to business travelers. Of the four species of malaria that affect humans, falciparum malaria is most likely to be lethal or disabling, especially the cerebral form. It is geographically dispersed throughout sub-Saharan Africa and southeast Asia, including New Guinea. (The others are vivax, which predominates in southern Asia and north-central South America, ovale, and malariae, both less common, and found in a patchy distribution where other malaria is endemic.) Measures for malaria prevention include local mosquito control (with the controversial issue of the proposed reintroduction of low-level treatment with DDT for this purpose), pesticide-impregnated mosquito netting, chemoprophylaxis (with particular reference to the high frequency of unacceptable side effects of mefloquine and the relative underutilization of doxycycline), mosquito-resistant clothing, and mosquito repellant (DEET).

Guidelines on the use of these modalities (always in combination) and current recommendations for malaria prevention should always be sought from the most up-to-date authoritative sources, such as Web sites of the Centers for Disease Control and Prevention and the World Health Organization, in preference to textbooks and other sources that are much less often updated.

A particularly hazardous situation occurs when workers invade unfamiliar ecosystems and encounter zoonotic (animal-borne) pathogens with a previously restricted range. An extreme example of occupational infection due to location occurred during the building of the Panama Canal. In the nineteenth century a pioneering French effort failed due to the high attack rate and mortality among workers of yellow fever, which had originally been introduced to the Americas from Africa and had become established in the rain forest. The success of the subsequent American effort was largely credited to mission-specific public health research, leading to the recognition of the role of the *Anopheles* mosquito in transmission, followed by practical and effective measures for mosquito abatement. A more recent, catastrophic example may be the emergence of HIV, the most significant among retroviruses that may have emerged in this manner, from primate hosts in Africa to disseminate into a global epidemic.

### **Community-Acquired Infection**

Although demonstrating nothing like the mortality and social impact of malaria and HIV, common infections circulating in the community naturally appear in the workplace as well and profoundly affect job performance. Absence and presenteeism (reduced productivity in a worker who is present but not functioning well) is highest from upper respiratory disorders and influenza, which are both very common disorders and easily transmitted in the workplace. Because of their frequency and transmissibility, they are the most important community-acquired illnesses introduced into the workplace and so will be used as examples.

As a general preventive measure, the most important intervention to prevent upper respiratory tract infections is clearly hand washing. Outside of healthcare, however, few industries promote this or make it a central theme in health promotion programs.

Control of upper respiratory disorders in the workplace requires measures to prevent the ill worker from introducing and transmitting the disease. This may be achieved by policies encouraging the worker to stay home during an acute illness. From the employer's point of view this may seem to be a mixed message, in that absence is being encouraged, but the risk of mass presenteeism caused by many workers struggling through the day with symptoms outweighs the temporary absence of one employee who may transmit the disease. Another approach to employers who are equipped for it is remote work from home, sometimes called "telecommuting," if the employee is well enough and doing office or creative work. This is not popular among managers, who are usually concerned that it will set a precedent for other employees wishing to work from home at other times and for their own reasons and because productivity is seen as more difficult to measure than for work at the office.

Work at home may also be an emergency measure in the event of widespread epidemic and transmissible disease, as proposed for pandemic influenza. This strategy, combined with canceling nonessential meetings or gatherings, discouraging employees from assembling into groups, and limiting face-to-face interactions is known as "social distancing" and is a prominent element in pandemic control contingency planning.

Immunization programs for employees have grown in popularity in recent years as their impact on workforce absence during influenza outbreaks has been demonstrated. This is a service that benefits individual workers and their families as well as the workforce overall and the community, while maintaining productivity and reducing costs. Compliance rates for employees vary, of course, and at least in hospital settings it is disturbing to note that the lowest compliance rates are generally found among

nurses and physicians. This demonstrates that awareness and cognitive knowledge do not automatically translate into health-conscious behavior.

## RESOURCES

- Ad Hoc Committee [on Update of 1986 Criteria for the Diagnosis of Nonmalignant Asbestos-Related Disease: Guidotti TL, Miller A, Christiani D, Wagner G, Balmes J, Harber P, Brodkin CA, Rom W, Hillerdal G, Harbut M, Green FHY]. Diagnosis and initial management of nonmalignant diseases related to asbestos. *Am J Respir Crit Care Med.* 2004;170:691–715. (Source of Figures 17.14, 17.15, 17.16, and 17.17.)
- Craighead JE. *Pathology of Environmental and Occupational Disease*. St Louis: Mosby; 1995.
- Harber P, Schenker M, Balmes J. *Occupational and Environmental Respiratory Diseases*. St Louis: Mosby; 1995.
- Quint J, Beckett WS, Campelman L, Sutton P, Prudhomme J, Flaherty J, Harrison R, Cowan B, Kreutzer R. Primary prevention of occupational asthma: identifying and controlling exposures to asthma-causing agents. *Am J Ind Med.* 2008;51:477–491.
- Sataloff RT, Sataloff J. *Occupational Hearing Loss*. 3rd ed. Boca Raton, FL: Taylor and Francis; 2006.
- Sullivan JB Jr, Krieger GR. *Clinical Environmental Health and Toxic Exposures*. Philadelphia: Lippincott Williams & Wilkins; 2001.
- Wright WE, ed. *Couturier's Occupational and Environmental Infectious Diseases*. Beverley Farms, MA: OEM Press; 2008.

## NOTEWORTHY READINGS

- Adams R. *Occupational Dermatology*. New York: Grune and Stratton; 1982.
- Adams R. Occupational skin disease. *Occup Med.* 1986;1(2):199–356.
- Blanc P, et al. An experimental model of metal fume fever. *Ann Intern Med.* 1991;114:930–936.
- Brain JD, Beck BD, Warren AJ, Shaileh RA. *Variations in Susceptibility to Inhaled Pollutants*. Baltimore: The Johns Hopkins University Press; 1988.
- Brooks SM, Kalica AR. NHLBI Workshop summary: strategies for elucidating the relationship between occupational exposures and chronic air-flow obstruction. *Am Rev Respir Dis.* 1987;135:268–273.
- Brooks SM, Weiss MA, Bernstein IL. Reactive airways dysfunction syndrome (RADS): persistent asthma syndrome after high level irritant exposures. *Chest.* 1985;88:376–384.

- Chan-Yeung M. Occupational asthma. *Chest*. 1990;95(suppl 5):1485–1615.
- Chan-Yeung M, Lam S. Occupational asthma. *Am Rev Respir Dis*. 1986;133:686–703.
- Cockcroft DW. Occupational asthma. *Ann Allergy*. 1990;65:169–175.
- Cordasco E Sr, Demeter SI, Zenz C. *Environmental Respiratory Diseases*. New York: Van Nostrand; 1995.
- Cullen MR. Clinical surveillance and management of occupational asthma: tertiary prevention by the primary OEM physician. *Chest*. 1990;98(suppl 5):196S–201S.
- Enright PL, Hyatt RE. *Office Spirometry*. Philadelphia: Lea and Febiger; 1987.
- Guidotti TL, Abraham JL. Occupational lung diseases. *Am Fam Phys*. 1984;29:169–176. (Source for Figures 17.7, 17.10, 17.23, and 17.25.)
- Guidotti TL, Miller A, Christiani DC, Wagner G, Blames J, Harber P, Brodkin CA, Rom W, Hillerdal G, Harbut M, Green FHY. Nonmalignant asbestos-related disease: diagnosis and early management. *Clin Pulm Med*. 2007;14(2):82–92. [Contains material following up on the revised ATS criteria.]
- Guidotti TL, Novak RE. Hearing conservation and occupational exposure to noise. *Am Fam Phys* 1983;28(4):181–186. (Source for Figure 17.24.)
- Hogan DJ, Dannaker CJ, Maibach HI. The prognosis of contact dermatitis. *J Am Acad Dermatol*. 1990;23:300–307.
- King TE, Jr. Bronchiolitis obliterans. *Lung*. 1989;167:69–93.
- Kusaka Y, Hering KG, Parker JE. *International Classification of HRCT for Occupational and Environmental Respiratory Diseases*. Berlin: Springer; 2005.
- Maibach HI, ed. *Occupational and Industrial Dermatology*. Chicago: Year Book; 1987. [Classic. Now hard to find.]
- Mathias CG. Prevention of occupational contact dermatitis. *J Am Acad Dermatol*. 1990;23:742–748.
- Mazurek JM, Wood JM. Asbestosis-related years of potential life lost before age 65 years: United States, 1968–2005. *MMWR*. December 12, 2008;57(49):1321–1325.
- Morgan WKC, Seaton A. *Occupational Lung Diseases*. 2nd ed. Philadelphia: WB Saunders; 1984.
- Perrin B, et al. Reassessment of the temporal patterns of bronchial obstruction after exposure to occupational sensitizing agents. *J Allerg Clin Immunol*. 1991;87:630–639.
- Salvaggio JE. The impact of allergy and immunology on our expanding industrial environment. *J Allerg Clin Immunol*. 1990;85:689–699.
- Weaver JH. Monitoring hearing loss at United Airlines. *Occup Saf Health*. 1981;50:35–39, 41.
- Weill H, Turner-Warwick M. *Occupational Lung Diseases: Research Approaches and Methods*. New York: Marcel Dekker; 1981.

*This page intentionally left blank*

# **18 CAPACITY FOR WORK**

This chapter deals with several important topics in occupational medicine that are connected in that they relate to the capacity of the worker to do the assigned work. These topics are:

- Disability rights
- Fitness for duty
- Certification of illness and medical leave
- Family and medical leave
- Absence
- Presenteeism
- Functional capacity evaluation
- Impairment and disability assessment
- Accommodation
- Drug screening

## **DISABILITY RIGHTS AND THE AMERICANS WITH DISABILITIES ACT**

Disability rights, as a movement, has deep roots in American, Canadian, and British law and society. The emphasis initially was on human custodial treatment, then in the eighteenth century on education and institutionalization into a separate society, then on “overcoming a handicap” and participating in mainstream life. As technology and medical treatment improved, assistive devices and rehabilitation supported the inclusion of disabled people as full participants in society. The achievements of exceptional athletes and the example of disabled veterans also demonstrated that a rigid distinction between the disabled and the whole is spurious. Capacity is a continuum, and disability is a mismatch in a particular context. Disability that has meaning with respect to work may mean little in terms of a social or family role. Modeled on the civil rights movement, disability rights advocates in the United States first demanded more responsive services and then broadened the issue to removing barriers to participation, culminating in 1990 with legislation that provided employment rights. Since then, a major theme of the disability rights movement has been parity with respect to mental health care and recognition of nonvisible disabilities.

In the United States, the Americans with Disabilities Act (ADA) of 1990 changed the landscape at the time with respect to perceptions of disability and work capacity. ADA has been a pervasive influence in shaping what is perceived to be appropriate with respect to work requirements and individual capacity. Occupational and environmental medicine (OEM) physicians with any part of their practice in the United States are advised to become intimately familiar with ADA, which was amended in 2008. (Canadian employment law is equal to or more stringent than ADA because of the antidiscrimination provisions in the basic law of the Charter of Human Rights and Freedoms, which overrides other legislation, the Canadian Human Rights Act of 1977, which applies to federally regulated enterprises, and the provincial human rights codes, which apply to everything

else. The major practical difference for employers is that human rights violations in Canada are heard by tribunals rather than being litigated in court. The British counterpart is the Disability Discrimination Act of 1995.)

With ADA, the emphasis shifted to the design of the job, identification of its essential requirements, and identification of reasonable accommodations that may be used to meet those requirements when a worker has a limited disability. ADA applies only in the United States but its provisions are consistent with Canadian law and precedent, and it represents a convenient framework for reviewing the fundamental principles of disability and work capacity.

Although ADA is written as a series of restrictions and prohibitions, it is actually legislation oriented to positive change that asserts a right of the disabled to fairness in hiring. It creates a framework within which the disabled applicant for a job can compete for jobs that he or she could perform and shapes the characteristics of the job to the capacity of the individual, within reasonable limits. Once hired, the worker who is able to demonstrate good work performance is promoted and treated on an equal basis with nondisabled workers. By requiring accommodation, so that persons who are able overall to do the job but need assistive devices or work modifications to do a part of their duties, ADA shifted the emphasis from finding a hire who fits perfectly into the slot to shaping the workplace with flexibility to fit the worker. This opens up employment prospects to many more people who can do the work. It also has a transformative effect on the workplace, because an adaptable workplace is generally a safer and more flexible workplace.

ADA prohibits discrimination in recruitment, hiring, advancement and promotions, job-related training, compensation, social activities, and employment benefits, including workers' compensation coverage and benefits, on the basis of disability. In order to ensure that discrimination will not take place in the hiring process, ADA restricts questions that can be asked about an applicant's disability before a job offer is made and does not permit medical evaluations before a job offer is extended. ADA requires

employers to make a “reasonable accommodation” for new hires who have limited disabilities, short of the point that accommodation would result in “undue hardship” for the employer.

Medical evaluation may only be conducted after a job offer, must be matched to the work requirements of the specific position, must be given consistently to all workers who are candidates for such a position, and must be kept confidential. The purpose of the medical evaluation is to determine whether the selected worker, who is already deemed qualified and has been offered the job, can do the job and do it safely, not whether they are in good health or at risk for a future health problem. No medical question may be asked of job applicants before an offer of employment has been made. Employers may ask current employees questions about their health and impairment only when there is a visible, concrete reason to believe that the worker cannot perform the job or poses a risk to self or others due to the condition.

Key words in ADA include a “covered entity,” which is generally an employer engaged in interstate commerce that has fifteen or more employees but may include employment agencies, labor organizations, trades, and joint labor-management committees. A “qualified individual” is a person who has a disability, visible or not. Drug abuse is excluded as a qualification. “Disability,” for purposes of the Act, is a “physical or mental impairment that substantially limits a major life activity.” What qualifies as a major life activity is decided by the Equal Employment Opportunity Commission. This provision led to a number of court challenges that confused interpretation of ADA, but in 2008 the Act was amended to clarify that major life activities are to be construed broadly, that “substantially limits” means “significantly or severely restricts,” and that a limitation does not disappear from consideration because a disabled worker manages in daily life with ameliorative measures, such as eyeglasses or a walker.

Other American legislation that may affect occupational health is the Architectural Barriers Act, which requires that federal facilities comply with standards for physical accessibility, the Rehabilitation Act, which prohibits discrimination on the basis of disability with

respect to employment and participation in federally sponsored programs, and the Air Carrier Access Act, which prohibits discrimination in scheduled air transportation serving airports in the United States and applies to accessibility to aircraft and assistance in boarding.

With respect to the broader world of environmental barriers and disability, ADA provides for barrier-free access in public accommodations and prohibits discrimination in services on the basis of disability by businesses that serve the public and on the part of state and local governments. Another title prohibits discrimination and requires accessibility in public transit and requires provision of “paratransit” (special services for the disabled who cannot manage public transit) where feasible. The Fair Housing Act, as amended in 1988, prohibits discrimination in housing on many grounds, including disability, and may require owners to provide access for persons with disabilities (in new multifamily housing with four or more units), to allow reasonable accommodations (such as allowing the tenant to modify doors to the width of wheelchairs, at the tenant’s expense), and making reasonable exceptions in policies that exclude persons with disabilities (such as allowing guide dogs despite a policy of no pets). The Telecommunications Act, as amended in 1996, requires manufacturers and service providers to provide equipment and services that can be used by persons with disabilities. A title of ADA also requires “common carriers” (telecommunications utilities) to provide services accessible to the deaf.

## **FITNESS FOR DUTY**

The OEM physician conducts, reviews, and designs fitness-for-duty (FFD) evaluations and in many cases supervises the FFD evaluations conducted by other physicians. FFD evaluations are a family of medical services that are specific to occupational medicine and involve (1) assuring that the worker or new hire is physically able to do the work assigned, (2) assuring that the worker or new hire can do so without presenting a safety risk to themselves or others, and (3) identifying any gap between the specific requirements of the job and the individual

capacity of the worker and determining the accommodation that may be required in any essential function that might enable the worker to do the job despite an impairment. These evaluations must be applied universally and cannot be conducted selectively on certain applicants and not others. They must be directly related to the specific job requirements and conducted by persons who are familiar with both the working conditions in that workplace and the limitations of currently available evaluation procedures. They must be kept completely confidential, and only judgments regarding fitness for duty, not diagnoses, should be communicated to management by the physician conducting the evaluation, unless the worker gives approval.

Fitness for duty implies that the worker has the capacity to do the essential duties of the assigned job and is not impaired in such a way that essential functions are impossible to perform. It implies no more than this and certainly does not require that the worker have no medical condition, nor does it imply a judgment regarding future capacity to function. Fitness for duty is a critical evaluation for OEM physicians to make and a major part of OEM practice.

Fitness for duty is often referred to as “fitness to work,” especially in the older literature. The terminology has changed because fitness to work implied a level of health or capacity for work in general, not limited to a specific job. Fitness for duty suggests fitness for a specific set of requirements that go with the job.

Historically, employers generally required their employees, especially new hires, to undergo medical evaluation (usually consisting of a general physical examination and basic laboratory tests, sometimes including a chest film) in order to determine whether they were healthy. These medical evaluations were often repeated on an annual basis as periodic health surveillance. The requirements for the “medical” or the “employment physical” were similar to a standard medical examination and bore little or no relationship to the job duties. The breakthrough in thinking about fitness for duty came in the realization, underlying ADA, that disability is not a condition of the individual but a mismatch between the specific capacity of the individual and the specific, essential requirements of the job and that a disability

in one setting may be irrelevant in another. In most cases, a relatively limited disability may also be easily overcome in the setting in which it is significant by a simple accommodation (for example, amplification for the hearing impaired). ADA provided a legal framework that realized a change in attitudes and culture regarding the disabled that had been evolving for decades. ADA focused on the essential duties of the job and turned the pre-placement evaluation into a structured analysis of the work assignment rather than an arbitrary test of the worker.

The primary objective of an FFD evaluation is to create a fair match in which the employee is fairly evaluated for his or her capacity to perform the assigned job in a technically satisfactory manner while the employer's rights are protected by ensuring that the individual will be able to perform the job as required and will not pose a safety threat to themselves or others. The FFD evaluation also has secondary benefits. It documents the employee's health status objectively so that future claims of injury may be dealt with on the basis of fact rather than supposition. It establishes a point of comparison so that should the employer fail to provide a healthy workplace, the lapse can be detected through medical monitoring and corrected accordingly. FFD evaluations cannot be used to speculate on future disability or prognosis and do not in any case accurately predict sickness absence in the future for an individual worker.

### **Designing the FFD Evaluation**

The FFD evaluation must be identical for all applicants for a given job description, matched against explicit job performance criteria, and treated confidentially.

Table 18.1 presents an example of how not to do it, a typical example of a regressive FFD evaluation of the type popular in the 1970s, incorporating irrelevant medical tests. It is presented here because such protocols are still in use and some employers still ask for them. They should be recognized by the OEM physician as an inappropriate service.

**Table 18.1.** An Inappropriate “Pre-employment” Protocol

This pre-placement evaluation is typical of those in use in the 1970s.

They are sometimes still encountered and reflect poorly on the employers who request them.

Medical history (may contain numerous irrelevant questions)

Physical examination (may include inappropriate and intrusive examinations)

Complete blood count

Blood biochemistry

Urinalysis

Chest x-ray

Low-back x-ray\*

Electrocardiogram

Stool guaiac for occult blood\*\*

\* Unethical because of excessive radiation exposure and low predictive value.

\*\* Often included as a personal health benefit.

A major limitation on FFD evaluations is the sensitivity and specificity of medical screening tests. (See Chapter 5.) The tests that physicians use for FFD evaluations are adapted from clinical tests used by physicians to assess people with illnesses. They were originally designed to assist in making a diagnosis in a person who presents to the physician as clearly ill. They are seldom very sensitive or specific for disease and are mostly irrelevant to work capacity.

When one applies these tests to a group of patients who are being investigated because they are likely to have the disease in question, the lack of specificity presents relatively little difficulty in determining predictive value. Knowing that an individual has a high likelihood of having the disease or impairment in the first place, the physician uses the test for guidance in selecting among alternative diagnoses that are highly probable (see Chapter 5). However, most groups of workers are relatively healthy and have a low prevalence of the conditions medical examination is testing for (historically, mostly tuberculosis and obvious clinical disease). FFD evaluations therefore have an inherent

problem with low predictive values of the screening tests used. It is not reasonable for an individual to be excluded from a well-paying job solely on the basis of a screening test with low predictive value.

Some tests are contraindicated because they carry an unacceptable hazard, as in the case of low-back x-rays. Low back x-rays do not accurately predict who will and who will not later develop back pain and carry an unacceptable radiation hazard; they are unethical for use in FFD evaluation. Even studies that withstand scrutiny are not terribly helpful in practice. Chest films are no longer useful because of the low prevalence of lung disease in the population today. Other tests are clearly irrelevant to work capacity: pelvic examinations were sometimes required for pre-employment medical examinations in the 1970s.

The findings obtained in a FFD evaluation must be compared against a set of criteria that are specific for the job assignment. For this reason FFD evaluations must be job specific, and requests to perform a FFD evaluation should always contain either a description of the job and the activities it requires or a set of validated health and performance criteria from an analysis of the tasks required by the job.

Formulation of specific performance standards and the derivation of medical criteria require a formal "job description." The job description is as detailed a description of the job as is required to identify the health and performance standards that should apply. The point of an FFD evaluation is to match an understanding of the job and the health requirements it entails with an understanding of the worker's capabilities and health, by comparing the health findings obtained from the evaluation to the working conditions as abstracted in a set of health standards for the job supplied to the examining physician. The physician is expected to be able to render a medical opinion regarding the fitness of that worker to work at a specific job.

An adequate job description useful to support an appropriate FFD evaluation should include at least the following:

- A description of the job function and performance requirements
- Location of work (i.e., special or hazardous work environment)

- Hours of work
- Occupational safety hazards
- Occupational health hazards
- Special or unusual requirements of the job
- Psychosocial demands of the job

Table 18.2 provides an illustration of an acceptable job description. This description was developed by the personnel office of a large public employer in California. The criteria for fitness for duty are implicit and can be easily inferred by a knowledgeable physician.

The description can then be used to decide on an appropriate FFD evaluation. The work mentioned above has resulted in sets of health standards that more or less dictate what tests should be conducted.

Obviously, there can be no justification for FFD evaluations for jobs that have no exceptional responsibilities or duties. There is no reasonable justification for performing elaborate medical evaluations on office workers or individuals whose jobs have no special physical requirements.

**Table 18.2.** Essential Features of the Pre-placement Evaluation

---

1. Scheduled *after* a job offer has been made, never a requirement for hire.
  2. Contingency for job offer to be final, not for eligibility for employment. “Pre-placement” is *not* a synonym for “pre-employment.”
  3. Specific to work requirements of the assigned job.
  4. May identify conditions that can be acceptable to employer with reasonable modifications in workplace.
  5. Routinely administered, in a consistent manner, to all workers assigned to the job, not applied selectively to those with visible or declared disabilities.
  6. Physician should never “approve” or “disapprove” employment. Qualification to be hired is a human resources decision, not a medical decision.
-

Whenever possible, the best evaluation procedure of all is to allow the employee to do the job on a trial or probationary basis and to assess performance, which is usually best evaluated by the immediate supervisor. If the individual can perform the job over a period of days and shows no sign of failing performance or a health problem, then the evidence of their capacity to do it should be clear. If the individual cannot do the job, however, this challenge test is much more sensitive than a clinical evaluation. The exceptions would be those jobs that present an unacceptable risk due to an intrinsic hazard or in which the safety of others may depend on the performance of the job by the worker (as in safety-sensitive positions). In these circumstances there is really no acceptable trial possible, and a test must be adopted.

### **Reporting Outcomes of FFD Evaluation**

Once the FFD evaluation is completed and the physician has compared the findings of the medical evaluation with the health standards appropriate for that particular job, an opinion is rendered regarding the worker's fitness for duty.

There are three possible outcomes that the physician can report:

- Fit
- Unfit
- Fit subject to work modification [specify]

These outcomes are in all cases specific to the job under consideration, not a general statement of the worker's capabilities or ability to work in any job.

“Fit” means that the individual is able to work in the assigned job without restriction; is not likely to pose a threat to self or others in performing his or her duties; and is not likely to suffer adverse health consequences in performance of normal duties.

“Unfit” means that the individual is not able to work in the duties assigned and is not likely to be able to do so despite reasonable accommodation or work modifications.

“Fit subject to work modification” means that the worker would be able to return to work if the job or working conditions were modified to accommodate limited impairment, which may be temporary or permanent.

In the past, it has been common for physicians to use the term, “light duty” with respect to return-to-work determinations. Employers do not find these designations to be useful or practical, as discussed below in the relevant subsection. It is better to use the modern terminology “fit subject to work modification” and to specify the type of modification that is required. Whether the work modification can be accommodated then becomes a management decision, not a medical one.

When an evaluation is conducted by a physician outside the employer’s occupational health service, the medical findings should not be reported to management or any representative of the employer other than the employer’s occupational health service or a designated occupational health physician or nurse. Confidential information should never be sent to nonmedical personnel without the patient’s explicit approval. The physician should not share confidential information on the employee with management. Only the medical evaluation regarding fitness for duty should be communicated to management.

### **Specific Types of FFD Evaluations**

Fitness for duty involves the following types of evaluations:

- Pre-placement
- Return to work
- Performance-related (medical inquiries)
- Job change
- Safety-sensitive positions (including U.S. Department of Transportation)

Each will be discussed in turn.

### ***Pre-placement Evaluations***

Pre-placement evaluations are conducted to determine the suitability of an applicant for a job to which they would be assigned. The job offer is made contingent on one's being fit for the specific job. The essential features of a pre-placement evaluation are listed in Table 18.3. Because of the provisions of equal opportunity legislation, the pre-placement evaluation has replaced the "pre-employment" evaluation.

Pre-employment evaluations were general screens for health status that were conducted prior to a job offer. Before ADA, the applicant

**Table 18.3. A Satisfactory Job Description**

---

#### MAINTENANCE, HEATING AND AIR CONDITIONING

##### MECHANIC

*Definition*

Under direction to perform skilled work in the repair and maintenance of heating, refrigeration and ventilation equipment; and to do related work as required.

*Distinguishing Characteristics*

The Lead Maintenance, Heating and Air Conditioning Mechanic is used to supervise the work of other skilled heating and air conditioning mechanics and perform skilled repair work personally.

*Examples of Duties*

Inspects, disassembles, repairs, maintains and services ventilation, air conditioning and heating systems and equipment; adjusts and installs valves, thermostats, fans and controllers; checks temperatures, pressures and differentials; lubricates, cleans and adjusts equipment; changes filters and strainers; flushes boilers, descales tubing, cuts and threads pipes; replaces valves, fans, motors, gaskets, filters and belts; uses various testing devices to locate defective parts; adjusts and installs switches, gauges, thermostats, valves, tubing and other parts as needed, makes pipe connections and general electrical repairs pertaining to equipment; schedules preventive maintenance; maintains logs and records on equipment repair or malfunctions; orders parts, maintains liaison with vendors; performs other maintenance tasks as assigned.

(Continued)

**Table 18.3. (Continued)**

*License*

Possession of an appropriate California operator's license issued by the State Department of Motor Vehicles.

*Desirable Qualifications*

*Knowledge:*

Tools, materials, methods and terminology used in the maintenance and repair of heating, air conditioning and ventilation equipment; The proper operation of heating, air conditioning and ventilation equipment.

*Skills:*

Trades level skill in the maintenance and repair of heating, refrigeration and ventilation equipment;

Estimate the scope and cost of work assignments and select necessary tools, equipment and materials to complete the job;

Use required tools and equipment skillfully and safely;

Work from sketches, drawings and blueprints;

Keep work records;

Understand and carry out oral and written instructions;

Establish and maintain cooperative working relationships;

Read and write at the level necessary for successful job performance.

*Training and Experience:*

Any combination of training or experience that could likely provide the required knowledge and skills is qualifying. A typical way to obtain the knowledge and skills would be one year of journey level heating and air conditioning repair experience.

for a job was often required to pass a "physical" before being hired, which put the physician in an untenable position of gatekeeper, deciding who was employable, based on criteria of little relevance to the work to be performed. Often these medical evaluations had little or nothing to do with the specific job to which the applicant was to be assigned. In the 1970s it was not unusual for clerical personnel to be required to pass the same pre-employment medical evaluations as production workers.

Pre-employment evaluations reflected health issues held over from the early twentieth century. The employers were seeking to ensure that: (1) the new hire not introduce a communicable disease into the work-force, which was a holdover from the days in which tuberculosis was prevalent in the general population; (2) the new hire was sufficiently able to do physical labor, which became irrelevant to most jobs in the modern workplace and in any case is not well assessed by a medical evaluation; (3) the new hire was healthy and unlikely to be absent or to become ill on the job, which was a holdover from a time when morbidity in the general population was much higher; and (4) the work-force was strong enough to sustain productivity under pressure, which was a holdover from attitudes toward fitness in the military. Screening examinations in the military had provided a familiar but misleading model for fitness evaluation because the demands on the soldier and requirements for military readiness do not apply to most jobs in civilian life. It should not be necessary for a person to be in ideal health and free of disability to be hired to do a job he or she is able to perform.

By the end of the twentieth century, many employers had dropped pre-employment evaluations because they were not cost effective and many physicians were questioning the increasingly questionable rationale for them. After ADA, pre-employment evaluations became legal liabilities because they implied that the decision to hire would be based on a medical screening test, as opposed to a pre-placement evaluation that evaluated capacity for the specific work assignment after the job offer was made and the worker was deemed otherwise qualified.

The difference between pre-placement and pre-employment evaluations is more than just a legality. In the former a medical evaluation was assumed to be a requirement for joining the organization and therefore constituted a barrier to entry. The concept of pre-employment evaluations put the physician in a position of “approving” or “disapproving” employment on the basis of often arbitrary medical screening processes.

A pre-placement evaluation, on the other hand, implies that the evaluation is specific to the working conditions and to health standards that are relevant to the job requirements. A pre-placement evaluation may identify conditions that are acceptable to the employer with

reasonable and relatively inexpensive modifications in the workplace. This concept was missing, however, from the earlier concept of a pre-employment evaluation, which emphasized screening out any physical abnormality and often placed modestly impaired individuals, who could do the job very well with slight modifications, at a decided and unfair disadvantage. The modern pre-placement evaluation specifies the role of the physician in assessing fitness for duty for the specific job and establishes a more objective set of medical criteria to be followed.

### ***Return to Work***

“Return-to-work” (RTW) evaluations are conducted after a worker has been off the job for a prolonged period and is thought to be ready to return to work. Here, the goal is to ensure that a worker who has demonstrated an ability to perform the job in the past is sufficiently recovered to perform safely and reliably again. Many injuries or illnesses involve deconditioning, muscle atrophy, persistent fatigue, reduced exercise tolerance, or new conditions that may interfere with job performance. Simply being off work for more than a few weeks changes behavioral patterns and habits and makes it difficult to return to the pace and organization of work immediately. Whether the health problem was an occupational or personal injury or illness, the RTW evaluation ensures that functional capacity has been restored to the extent that the worker can resume performing the assigned job or identifies an accommodation that will allow phased reintegration into the workforce.

Ideally, an RTW evaluation is scheduled at the point when the worker is nearing or at end of a period of rehabilitation, or following a short illness or recovery from injury. Physical therapists are often asked to perform RTW evaluations at the end of a course of treatment, before discharging the patient. At that time gaps can be identified, for example in the limits on the number of repetitions of a hand action or fatigue in working a full day. Accommodations can be specified, such as an assistive device or partial work hours.

Returning the worker to the job as early as it is safe to do so helps to condition the worker physically, to restore the habits and structure of work, and to reintegrate the worker socially with the workplace. The employer should gain from early and safe return to work because it reduces costs associated with the total time off and restores an experienced employee.

The practical problem of RTW evaluations is that many employers will not accept workers who are “less than 100 percent” in work capacity. This is in part a reflection of an attitude on the part of management that they expect 100 percent from all employees, but it is also a response to perceived liability, because workers who are recovering from an injury are thought to be at greater risk of a second injury until they are fully healed. Employers may also be responding to practical issues of labor management. Temporarily assigning returning employees to jobs that are perceived as easier or lighter disrupts the organization of work and may create resentments among other workers who feel that they are expected to “carry” or “cover for” the recovering worker. Many unions are unenthusiastic about the practice because it disrupts seniority for the lighter jobs and is perceived as pushing the worker back before he or she is fully recovered.

The terms “light duty” and “partially fit” are often used as recommendations to employers for RTW determinations, often to encourage early return to work. The terms mean that the worker could return to work other than their usual job, as long as that work is not demanding given this impairment. Unfortunately, such a term is meaningless without a description of what the worker can or cannot do. It implies that the worker can be reassigned to other duties easily, which is often not the case in a highly structured and lean workplace. “Light duty” is a sloppy term and should be replaced with “modified duty with [specify restrictions or accommodation].” Many employers have a policy against any employee returning to the job until they are able to do their usual job safely with no special consideration in job assignment. The reasons for this include fears of re-injury and of legal liability if the employee cannot do the job safely. Light duty or alternative jobs simply may not be available in the workplace. Assignment

of employees to other duties may also be constrained by contracts with unions or may complicate labor-management relations, since one groups of employees may complain if preferential treatment is being given to one worker.

RTW evaluations are similar in concept and design to pre-placement evaluations and are conducted in a similar manner, with equal attention to the requirements of the job assignment. The worker previously assigned to the job is assumed to know how to do it, and his or her own opinion about capacity to do the job is usually reliable.

Physicians sometimes find themselves under pressure to extend the leave period well beyond the time when the worker is fit to return to work. Physicians are often urged by the patient to let them stay off work, particularly if there is a holiday, weekend, or social event ahead. The pressure may be even greater if the physician also treats the patient's family, has many other patients who know the insistent patient, or has close ties with the community and will inevitably run across the patient and his or her close associates again and again. It may seem a small matter to bend on a seemingly administrative issue. However, such an act is neither good medicine nor good occupational health practice. It is very costly to the employer and further delays the return of the patient to a normal life, which is part of the recovery process. The patient should return to work when the patient becomes fit to work, definitely not before but also not long after.

### *Medical Inquiry and Performance-Related Evaluations*

Performance-related FFD evaluations are medical evaluations triggered by the observation, usually by a supervisor, of a decline in job performance, unexplained and unexpected considering the worker's past performance. The essential features of this type of FFD evaluation are outlined in Table 18.4. These evaluations may be suggested by supervisors who detect failing or substandard performance in an individual with an otherwise good work record and who may be concerned that a health problem is interfering with the worker's ability to do the job. Under ADA, these are called "medical inquiries"

**Table 18.4.** Example of a Set of Health Standards Based on Criteria Derived from Analysis of the Specific Job (described in Table 18.3).

Critical features
Working near operating equipment
Bending, lifting, pulling
Working in awkward or cramped quarters
Hand tools
Critical exposures and work characteristics
Ergonomic
Vibration
Solvents
Noise
Health standards
Hearing surveillance (audiogram)
No medical condition precluding working alone
No medical condition interfering with detailed, close-up work
Musculoskeletal strength, flexibility, dexterity

and must be directly related to job requirements and “consistent with business necessity.” Asking about medications, for example, is a highly sensitive medical inquiry.

The U.S. Equal Employment Opportunity Commission (EEOC) recognizes that not all questions about health are medical inquiries that are sensitive with respect to discrimination. Some questions are allowable based on visible evidence, direct relevance to work performance, and legitimate business reasons to know. These include the following:

- How are you feeling today?
- Can you perform your job?
- Are you using alcohol or taking illegal drugs?

Such questions are allowable if the employer “has a reasonable belief, based on objective evidence” that the employee has a

condition (whether the disorder or medication required to treat it) that either impairs him or her from an essential job function or poses a direct threat to the safety or health of that employee or others.

Referral for a medical evaluation can often make the encounter with the worker tense and anxious and puts an additional responsibility on the physician to be sensitive to psychological factors and to identify problems such as substance abuse. Under ADA, the employer and the OEM physician who is involved must confine themselves to the particular issues of job performance and safety. Broader inquiries into the worker's personal health are not appropriate. Once the worker discloses an impairment, the medical inquiry is over except insofar as an accommodation can be recognized. There is no legitimate reason to know the details of a personal illness or to obtain the full medical history. As with all medical information, the findings in the medical inquiry must be kept strictly confidential but so must information regarding the accommodation.

The employer should have clear policies for when an FFD evaluation might be suggested or directed, and the OEM physician should have a protocol or plan for evaluation that conforms to fitness for duty for the work assignment and would be applied consistently to all employees in the same situation. The model of an independent medical evaluation (IME) should be followed (see Chapters 15 and 23), with the physician making it clear that his or her investigation is limited to assessing the reason for the decline in performance and not for purposes of treatment. Follow-up and therapeutic intervention should be done with the worker's personal physician, so as not to confuse the roles of objective evaluator and treating physician.

Performance-initiated FFD evaluations should be done in conjunction with the usual procedures for an employee assistance program, if one is available. This opens the possibility of intervention and rehabilitation. It is important to maintain a relationship with the worker-patient that is cordial and appropriate to a medical encounter rather than a personnel action. The physician should emphasize the role of the FFD evaluation in maintaining a balance between the obligations and rights of the employee and the employer.

Performance-based FFD evaluations offer the opportunity to identify a previously unsuspected problem and to find a solution that benefits both the worker and the employer. For example, if there is an on-site occupational health service or an on-site general medical facility, it may be possible for the OEM physician, occupational health nurse, or primary care provider in an embedded facility to monitor and help manage chronic medical conditions during the workday, such as brittle diabetes, asthma, and hypertension. These are opportunities for health gains in patient care and in managing employee wellness (see Chapter 19). The OEM physician should recognize that assuming direct-care responsibilities may make it more difficult to maintain objectivity in evaluating work capacity, and the arrangement should be governed by an explicit policy or by a written contract between the worker/patient and the physician. There should be agreement on all sides that the OEM physician will take on this role, it should be limited to the time that the worker is physically present at the facility, and care should be directed by the worker's personal physician with the OEM physician in a supportive role. Such arrangements are likely to become more common as on-site medical clinics become more common in large retail and employer facilities. Such arrangements obviously must be voluntary on the part of the worker, who may not wish to share personal medical information with another physician.

On the other hand, performance-based FFD evaluations are easily abused. Performance-initiated FFD evaluations should never be a covert form of discipline, but in the past they have often been used by supervisors and human resources departments to find a way to remove unwanted or troublesome employees. Requiring a medical evaluation may also be used as a threat by management to prompt the resignation of an employee who does not wish a health condition or a drug or alcohol problem to be revealed. The OEM physician should be aware of this and avoid getting in the middle by asserting professional independence and going by the book. Employees missing a certain number of days of work should not be automatically sent for review to the physician. The physician should

never, ever be expected to give the employee “a good talking to.” The OEM physician is not an enforcer and is not there to do the work of the human resources department. Being used in this way discredits the physician and undermines trust in the professionalism of the occupational health service on the part of both the employee and the employer.

Given the many sensitivities and pitfalls associated with ADA violations, restraint is advised. The OEM physician, when asked, should advise the employer that it may not be in the employer’s best interest to make a medical inquiry, especially if an accommodation can be made without knowing more about the worker’s condition. The employer does not need to know the diagnosis or details of a worker’s impairment and is better protected when management does not have this information. Policies and mechanisms to ensure confidentiality must be in place. There should always be a legitimate business reason or a medical inquiry is not appropriate.

It would be very wise to consult legal counsel experienced in ADA and EEOC case law before making any medical inquiry. Case law involving ADA and the EEOC is developing rapidly compared to other areas of law.

### ***Job Change–Initiated FFD Evaluations***

FFD evaluations initiated by a change in work assignment alone are rare but by the logic of ADA should not be. If a worker moves from one job with a given set of requirements and demands another job with very different requirements, an FFD evaluation ought to be provided just as if it were a pre-placement evaluation for a new hire. To do otherwise is to contradict the relationship between the initial evaluation and work requirements. However, there is no logical contradiction if the work requirements of the second job are incorporated into the evaluation of the first job or are less stringent.

The most common exception to this is that new jobs requiring the worker to use personal respiratory protection when they have not done so before always require an evaluation, as described in

Chapter 14. The evaluation incorporates tests for capacity to use respirators, which is an essential part of such jobs.

### ***Safety-Sensitive Positions***

Safety-sensitive jobs still require rigorous FFD evaluation. They are required in cases in which the physiological demands of the job are extreme or the consequences of inability to perform the job are intolerable. For example, the fitness and medical criteria for airline pilots are quite rigid. Deviations from them immediately ground the pilot. Similarly stringent FFD evaluations are required for divers, police, firefighters, and transportation workers in safety-sensitive jobs (such as drivers, railroad engineers, air traffic controllers, and marine pilots—merchant marine seafarers have a FFD that is less rigorous). In each case, the FFD evaluations are designed to protect the individual worker but also the public and other workers from harm. The FFD evaluations and the individual programs of which they are a part are highly specific and detailed, and it is beyond the scope of this book to provide more than a few examples. The most important in practice are those related to surface transportation and to aviation. Any OEM physician conducting these evaluations must train and achieve certification to do so.

Most OEM physicians, even in Canada, will at some time have some contact with the special evaluations of the U.S. Department of Transportation (DOT). (The United States Coast Guard, formerly part of DOT but now part of the Department of Homeland Security, handles FFD of mariners.) Many will qualify in order to conduct them, as they are one of most common services OEM physicians provide.

The Federal Motor Carrier Safety Administration (FMCSA) of DOT requires medical certification for up to two years of truck and bus drivers (“commercial drivers”), with exclusionary criteria for disabling conditions that may result in inability to control the vehicle. The Commercial Drivers Medical Examination (CDME), for example, requires visual acuity of 20/40 on the Snellen chart (corrected) and color vision for traffic signals. There should be no hearing deficit at more than 40 dB at

specified frequencies. The exclusions from certification include cardiovascular and respiratory disorders that carry a risk of loss of consciousness, insulin-dependent diabetes, epilepsy, current alcoholism, and a variety of conditions that would affect coordination and control of the extremities while driving. However, drivers who can demonstrate that the conditions are under control or have been successfully treated can apply for an exemption. Some of these conditions, when stable, require re-evaluation every year. Drug screening is mandatory, and a positive test excludes the applicant from driving. The certifying physician can make one of three decisions: certify, "time-limit" (certify for less than two years), or disqualify. There are many nuances in the CDME and much conferring among physicians who do this work to achieve consistency and fairness.

Obviously, the decision in a DOT is critical to the driver's future and the public safety. The essential evaluation may require follow-up of specific medical problems, such as hypertension, and there is some room for interpretation. In order to facilitate fair and consistent evaluations, DOT recognizes certification by examination of "medical examiners" who have undergone a training program but does not limit itself to any one training program or certification program. Certified medical examiners are entered into the FMCSA National Registry of Certified Medical Examiners, which is accessible to drivers. FMCSA is advised by an advisory network of medical expert panels.

Another safety-sensitive FFD evaluation is the Aviation Medical Examiner program of the Federal Aviation Administration (FAA) in the United States. The FAA qualifies pilots by classes: First class is airline transport pilots. Second class is commercial pilots. Third class is private pilots. Like the CDME, the FAA has standards for vision and hearing and many disqualifying conditions, most relating to potential causes of sudden incapacity or misjudgment. In order to retain high standards and consistency, the FAA manages a training and certification program to qualify physicians as Aviation Medical Examiners. Transport Canada has a similar program for Canadian Aviation Medical Examiners.

Safety-sensitive jobs in the private sector must be identified in advance. Medical qualifications must be justified as essential to the position, and a periodic medical evaluation must be a requirement of

the job. An example would be mine rescue personnel, who operate at maximum levels of exertion during an emergency and must therefore train vigorously. Their own lives as well as the lives of those they are trying to rescue depend on their capacity to work in unpredictable extreme conditions. Such a position would justify stringent medical evaluation on a regular basis.

## **CERTIFICATION OF PERSONAL HEALTH PROBLEMS**

Assessing a worker's ability to work as a result of personal health problems is usually a function of clinicians in the community, in the form of the familiar "note from the doctor." Certification of illness should be accepted as a medical function requiring conscientious management. This function places considerable responsibility on the physician, who has rarely had any training in how to do it and usually perceives it as a bother. The physician should never certify an illness that was not observed or sign a "blank check" when an employee presents claiming to have been ill for several days but not having sought medical attention. Unfortunately, this commitment is rarely taken seriously, and physicians continue to treat medical certification very casually.

Some employers ask the OEM physician to certify time off work as resulting from a medical or health condition. Certification of illness should either be performed with a knowledge of the employee's health problem or conducted as an FFD evaluation. In other words, the physician should not be placed in a position of having to write a note that states "certified off work because of influenza for last two weeks" but, more realistically, should be able to write "medically certified as fit to return to usual work after reported illness of two weeks." For the OEM physician, certification is actually an opportunity to conduct a less structured variation of the FFD evaluation.

For occupational physicians approving return to work and certifying time off work are a clear commitment to understanding the worker's personal health problems. If a worker's absence is likely to be prolonged or to be repeated, the physician should consult with

management to see if work modification or accommodation could be introduced to facilitate recovery or if some medical management could be undertaken by the occupational health service to ease the transition back to usual employment.

In fairness to the practicing physician, patients frequently ask for certification or a “note from the doctor” after the fact or for complaints that are impossible to confirm or deny. There is no excuse, however, for knowingly acquiescing to a false certification or to pad the amount of time for reasons of personal convenience. One way to deal with this is to treat it as an implicit FFD evaluation rather than a certification of an illness the physician has not seen.

## **FAMILY AND MEDICAL LEAVE**

In the United States, the Family and Medical Leave Act of 1993 (FMLA) sets the minimum allowable absence time that covered employers, generally those with more than fifty employees within seventy-five miles (not necessarily in one place), must provide if required to employees who qualify, generally those who have worked more than a year or a total of 1,250 hours. FMLA was under review by the U.S. Department of Labor at the time of this writing and may be amended by Congress in the near future, so readers are advised to seek current information.

Under FMLA, the employers with must grant eligible employees up to twelve work weeks of unpaid leave during any twelve-month period for the following purposes:

- Birth and care of a newborn child
- Placement with the employee of a son or daughter for adoption or foster care
- Care of an immediate family member (spouse, child, or parent) with a serious medical condition
- Medical leave when the employee is unable to work because of a serious health condition

Paid annual leave or sick leave may run concurrently with FMLA, providing the employee with income, or FMLA may be taken as unpaid leave with sick leave preserved. FMLA leave may be taken as a bloc or as intermittent leave for treatment. Employers are free to provide more generous policies and benefits, as long as the provisions of FMLA are present in full.

FMLA has functioned remarkably smoothly since it was introduced. Several problems have been noted with FMLA to date, notwithstanding the value of the act's provisions.

One problem has been that the burden of adapting to leave has fallen hardest on employers who work on tight schedules, with small numbers of skilled workers, in less flexible workplaces. That tends to describe many new start-ups or high-technology firms.

Another problem has been the definition of "serious health problem," which is sometimes disputed. For example, a qualified worker recently asked his employer for four days of medical leave under FMLA to obtain and to recover from intraocular lens placement for simple vision correction, not for cataract; this previously specialized procedure is coming to be used much like Lasik® for refraction. Although the case has not been arbitrated, it raises many questions. It would seem that this procedure would not be covered by FMLA because the underlying problem is not a "serious medical condition." Under FMLA, the employer has had the option of requiring the employee to get a second opinion from a physician of the employer's choosing, but could not name a physician who "contracts" with or regularly provides care for the employer; a third, mutually agreed-upon physician can break the tie. That provision excludes many OEM physicians. This is unfortunate because OEM physicians know about the workplace, possible accommodations, disability management, and fitness for duty and could assist in reentry.

A third major problem has been that time off for medical treatment is often not scheduled in advance with the employer, although routine medical appointments should be known well in advance. The employer then has to cover for unexpected absences.

The short duration of family leave provided by FMLA places a particular burden on nursing mothers who, after twelve weeks, must rely on employer policies regarding lactation breaks and the availability of suitable locations to express milk unless child care is on site and readily accessible. Breast feeding is a public health priority and a bond between mother and infant. Evidence suggests that exclusive breast feeding, optimally for six months, is associated with a reduced frequency of gastrointestinal disorders in the child in developed countries and reduced rates of Type I diabetes, and longer duration of breastfeeding appears to be associated with better health status for both mother and child. Breast feeding is associated with reduced infection rates and infant mortality in developing countries and in disadvantaged communities. However, return to work places nursing mothers who need to work but would prefer to nurse their child in a bind. This bind is only partially resolved by employer policies regarding lactation breaks and rooms.

FMLA has had a positive impact on worker's lives and has made employment more flexible and responsive to the demands of life. However, it must be noted that other countries, including Canada and members of the European Union, provide much more generous benefits for workers in need. In those countries, family and medical leave is part of the social contract between workers and employers. Seen from a global perspective, FMLA balances human needs against economic productivity in the United States, but it is nothing special.

## **ABSENCE**

Absence from work is a difficult management problem in industry, since it introduces considerable uncertainty into the scheduling of work and of staff assignments, is an important cause of lost productivity, and is a very common precipitating cause of labor-management conflict and misunderstanding. Because absence from work is viewed very differently depending on one's cultural background, the issue is laden with prejudice, misperceptions, and hidden hostility. There is no other area of working life where expectations of the work ethic are

more likely to conflict with notions regarding personal freedom and flexibility and where abuse so directly undermines the relationship between manager and worker.

Because absence triggers emotional responses and attempts to control unauthorized absence often makes matters worse, human resources personnel often persuade management to take a more “objective” approach by treating the problem as one of validation and documentation. Much of the total absence experience in industry is because of sickness and even more is claimed to be sickness. From management’s point of view, therefore, it may be only logical that absence be dealt with through medical means by requiring employee screening to identify workers likely to be frequently absent, medical certification of claimed illness, and investigation of individual cases. Some go so far as to expect their occupational health service, if they have one in-house, to monitor absence directly as a principal responsibility. Unfortunately, such responsibilities work no better than other tactics and have the potential to destroy the effectiveness of the occupational health service overall by turning it into the attendance enforcer.

### Absence Management

The transfer of responsibility for absence monitoring and control from the personnel or human resources department to the occupational health service converts an administrative problem into a medical and programmatic nightmare. It succeeds in getting the human resources department off the hook but presents the occupational health service with a fundamental inconsistency that workers are quick to see and some managers are quick to exploit further. When workers view the occupational health service as the means by which management reviews their attendance and singles them out for discipline, cooperation and goodwill promptly evaporate and are seldom if ever regained. The physician in such a position becomes labeled as just another management functionary, and medical judgments from the occupational health service are

viewed as untrustworthy and prejudicial to the worker. Having taken the occupational health service this far down the road, insensitive managers may then press even further for the physician to become a “team player” siding with management for violations of confidentiality.

Although the occupational health service should never accept monitoring and control of absence as an operational responsibility, there are many ways in which the occupational health service can have a positive effect on the absence experience of an employer to the mutual benefit of employer and worker. These include evaluating employees who are referred either voluntarily or individually by their supervisors for frequent or prolonged absence due to apparent health problems. This evaluation is an extension of the medical inquiry concept. A private, confidential interview and, if the worker allows, examination may uncover a treatable illness, may suggest a means by which the patient could be treated at work conveniently and without disruption to the work schedule, may suggest a minor modification in the work environment or in responsibilities that would allow the worker to stay on the job, or may identify a worker with emotional or substance dependency problems that might benefit from referral to an employee assistance programs.

In all such cases, the emphasis should be on fitness for duty and the well-being of the worker, not on policing compliance with the policy on sick leave. The occupational physician is primarily concerned with work-related health problems, but a question of at least as great of significance in practice is the assessment of a patient's ability to work in the presence of a personal health problem. There are circumstances in which a personal health problem is optimally managed and monitored at the worksite, as has been shown for hypertension, or in which the occasional attention of the OEM physician may prevent loss of work time and improve the quality of life for the worker, as in the management of brittle diabetes. Such individual patient management at the worksite should be considered complementary to the management of the worker's personal physician, providing the patient/worker with the best

opportunity to remain on the job while insuring that health care needs are met.

Health promotion programs in the workplace may reduce absence in various ways.

### **Absence Monitoring**

There are only two ways to keep track of absence: to record attendance at work or to record incidents of absence, assuming attendance unless notified otherwise. One is the time clock and the other keeps track more or less by voluntary notification or observation by supervisors.

For a variety of reasons that have as much to do with assumptions regarding social class and behavior as with flexibility in work, attendance recording is normally used for blue-collar and manual labor occupations and some office workers in automated or data-entry jobs in which wages are calculated by the hour and productivity is measured by quantitative standards or on a group basis. The implicit assumption is that without close tracking, the worker would not reliably show up to do the work and would abuse sick leave.

Absence recording, on the other hand, is usual for professional, white-collar, and most office workers who are paid on a salary basis and whose productivity is judged primarily by standards of individual performance. At higher levels of responsibility, the freedom to take time off whenever one's schedule permits is tantamount to an executive perquisite, but the socialization process of executives is such that while the option is provided, it is not expected to be used, let alone abused.

Absence is usually measured in terms of days, since this is the most convenient and universal unit to record. There are other measurements, however. Hours or shifts lost both reflect loss of productivity more directly, although it should be kept in mind that presence at the worksite may represent the opportunity to be productive better than actual productivity. For example, a worker in an academic or creative field may find inspiration at any time on or off the job and may

spend relatively little time actually producing a tangible product but that product may have high value, while an office worker with ambiguous or poorly defined responsibilities in a large and complex organization may sit at a desk daily for years with no substantial output. For most jobs, however, time put into the task bears some relationship to productivity.

The frequency of “spells” (the term of art used in the older literature), or identifiable episodes of absence, may reveal a pattern suggestive of a medical cause. For example, binge drinkers not uncommonly lose a day or so preceding or following weekends or holidays. Their pattern of absence may be frequent but not necessarily excessive in terms of lost days. The severity of the problem may not be reflected in the numerical count of lost work days. An individual with a severe, chronic illness may lose many days, but only have a few episodes of illness per year.

Once counted, absences can be compared and monitored by the calculation of summary measures, each with their own limitations:

- Frequency rates reflect the number of episodes of absence per year, for a person or the average for a group; they may include all episodes of absence, short-term absence only, or only absence lasting more than a few days depending on the purpose for which the statistic will be used.
- Severity rates are based on the number of days (or shifts or hours) lost per year, either for an individual worker or the average for a group; these are sometimes expressed as the percentage of total working time lost due to absence.
- Prevalence rates are calculations of the percentage of employees absent on a particular day; unlike the other two types of measurements, prevalence rates cannot be calculated for individuals and are not meaningful beyond one point in time, but they are useful in making estimates or evaluations of manpower needs and are commonly used in departments of human resources. (A common benchmark for general industry is 2 to 3 percent.)

In practice, employers who track absence use many different systems to calculate and to express absence, and there is little consistency among companies or between private corporations and public agencies except for the crude measurement of days lost. Comparisons among organizations are difficult, and no absolute standards exist by which to compare an employee's experience with others in their industry or community. One should not confuse measures of absence with epidemiologic measures of the frequency of illness such as morbidity, incidence, point prevalence, and period prevalence, each of which has precisely defined meanings. Absence measurement is just that and cannot be readily converted into measurements of morbidity without additional information, which are rarely available in practice.

Absence rates vary greatly from place to place, employer to employer, industry to industry, and with socioeconomic factors such as the state of the economy and available sick leave benefits. There is no generally accepted "normal" absence rate that can be used as a standard of comparison. In general, however, absence rates have slowly risen over the last thirty years; this is true throughout the developed world.

Absence can be categorized in many ways, but the most satisfactory, simple system recognizes five basic categories:

- Sickness absence, defined as absence attributed to incapacity due to illness or injury that is not work related.
- Personal leave, defined as absence for personal reasons having nothing directly to do with incapacity.
- "Time-lost" occupational injuries or, much less commonly, illnesses. By definition, these are reportable under workers' compensation and are therefore closely monitored.
- Pregnancy and child-care leave. By its nature, this is planned and preauthorized under the employer's policy, which is regulated by the Family and Medical Leave Act in the United States.
- Leave to care for a dependent.

- Bereavement.
- Partial absence (late arrival, early departure, temporary absence).

Each of these has further subcategories and nuances of measurement, as suggested in Table 18.5. Sickness absence and personal leave will be discussed in greater detail. These categories are useful on an aggregate basis for administrative purposes, in keeping track of reasons for absence. It is not appropriate to track these patterns for individuals.

A further problem with monitoring absence is that it is often to the worker's or the employer's benefit to misclassify absence. A worker may call in sick rather than take allowed vacation time in

**Table 18.5.** Classification of Absence

- 
- Sickness absence (up to maximum allowed by sick leave policy)
    - Certified (by physician)
    - Uncertified (up to 3–5 days usually permitted), or “self-certified”
    - Prior authorization (e.g., for a physician's appointment)
  - Personal leave (with or without pay, with or without authorization)
    - Personal business
    - Family illness
    - Bereavement
    - Jury duty
  - Occupational injuries and illness (“time lost”)
  - Family and medical leave
    - Antenatal and confinement
    - Postpartum and early child care
    - Paternal
    - Medical problem in family
    - Self-care
    - Care of dependent adult
  - Partial absence
    - Late arrival
    - Early departure
    - Prolonged absence from post
-

order to preserve paid sick leave or to take the day at a more convenient or spontaneous time. An employer may count a real illness that is inadequately documented against an employee's vacation time in order to avoid paying wages for the time off. Abuses of the system occur in every industry and in almost every workplace; a perfect system has yet to be devised.

The primary classification problem from the point of view of industry is differentiating sickness absence from unauthorized personal leave. This is typically done by requiring a physician's "certification" of the illness, as discussed in the previous section. Certification is intended to be a check on the employee's declaration of illness, but leads to its own problems and complications.

Historically, women are absent from work, on average, at about twice the rate of men if pregnancy is not counted. The reason is not that women use health services more often than men, although this is true, but that women generally have had greater domestic responsibilities and have been more likely to be employed in lower-status occupations in which time keeping is rigidly enforced and hours are inflexible. Women do not seem to have more severe illnesses on average than men and are less likely to work in hazardous occupations.

Smokers are, in general, absent more often with minor illnesses than nonsmokers. Abusers of alcohol not infrequently are absent sporadically, often on days following weekends and holidays and often without communication or notice. Despite these general trends, there is no accurate way of predicting absence for individual workers. Those with chronic illnesses are not uncommonly among the most reliable because they are committed and have learned to adapt. Normal, healthy workers may be absent for reasons having little to do with health or other characteristics that can be assessed in advance. The only predictor of absence in the individual case is the worker's individual history of absence. Even then, a person's pattern can often change, either for the better or for the worse.

Control of absence has been a major challenge to human resource managers in all industries. Few universally successful strategies have been identified, but a few observations have emerged. Personal leave,

including unauthorized personal leave masquerading as sickness absence, may be reduced by introducing flexible hours or permitting a certain number of days off per month or year to attend to personal business, subject to prior notification. This reduces the disruption caused by the need to attend to pressing personal errands or appointments during regular working hours and reduces the incentive to claim paid sick days. Institution of a sensible sick leave policy removes abuses engendered by the desire of some workers to "get back" what they may feel is owed them. A policy of requiring medical certification only after three days absence, as long as the worker phones to inform the supervisor daily, shows some degree of trust while minimizing the opportunity for flagrant abuse. The most effective control measure, however, seems to be an overall improvement in attitude and relations between labor and management. Absence is often affected by personal attitudes toward the employer, personal identification with the performance work and objectives of the group, and personal attitudes of self-respect. A general sense of personal satisfaction and of identification with the success and objectives of the employer may remove the temptation for abuse while motivating workers who are only somewhat inconvenienced by minor complaints to come to work.

Absence is determined in some part by attitude. Persons who have similar levels of illness and discomfort may vary considerably in their adoption of "sick" behavior, with one stoically carrying on and another taking to bed. Some of this difference is cultural, some is due to family attitudes and upbringing, and some is due to circumstances such as other stresses in the worker's life and individual patterns of strength and weakness. There are always a few individuals in any large group who are immature and who seize on any inconvenience as an excuse to avoid work; such people are a small minority, however. The majority see work as a responsibility to be met with varying degrees of commitment and use their working environment as an opportunity to make friends, to engage in social interaction, and to develop a social support system. Such workers may even come to work when they should stay home, out of identification with their group.

Finally, the alert physician whether in a corporate setting or in community-based practice, may spot a pattern in absence from work that leads to the identification of a health problem. One should be particularly alert to the following:

- Alcohol or substance abuse
- Depression
- Incomplete recovery, followed by subsequent absence or re-injury
- Covert illness, such as dementia, associated with inability to do or sustain the work
- Stress and emotional distress, whether personal or job-related, as in the case of “burnout”
- Abuse of sick leave, which may predict poor compliance or lack of cooperation

Control of absence is not a medical responsibility. Ultimately it is a performance issue that must be dealt with by human resources. The physician should never be placed in a position of monitoring absence or of being the “police” for attendance on the part of worker. Adopting this function puts the physician in an adversarial relationship with workers and interferes fatally with the worker’s perception of the physician’s perceived objectivity and fairness. The physician can, however, support the worker with a health issue and point the way to a reasonable solution.

The occupational physician can play a constructive role in evaluating individuals with repeated, unexplained, or suspicious absences for health problems that may require treatment, work modification, or close monitoring at work. Repeated spells of absence should be considered as possible indicators of a serious health problem. In cases in which repeated absence or failing performance indicates a problem, the physician may play a positive role by detecting the health problem before its effects are irreversible, before it interferes with the life and well-being of the employee and deprives the employer of a valuable worker.

## FUNCTIONAL CAPACITY EVALUATION

“Functional capacity evaluation” (or “assessment,” sometimes called “physical capacities evaluation”) is a comprehensive approach that arose in the 1980s from vocational and occupational rehabilitation and which incorporates insights from biomechanics, kinesiology, and exercise physiology. A typical functional capacity evaluation would test up to twenty different physical characteristics (such as the maximal strength of a worker, flexibility, agility, stamina, range of motion, hand-eye coordination, and so forth) and compare them against norms for the population and requirements of the assigned job. The objective is usually to profile the entire range of capabilities of a new applicant or of an injured worker before return to work.

Functional capacity evaluations have their highest application in assessing the maximal performance characteristics of the body. This makes the approach useful in situations in which it is useful to measure peak performance, such as athletics, military training, mine rescue, or research, or in public safety occupations in which peak or strained performance must be sustained in extreme conditions, most obviously firefighting. It could also be useful in determining what a person with multiple or interacting impairments can do, regardless of job assignment.

Work capacity varies greatly from person to person. The objective of functional capacity evaluation is to determine how much a person can do. The requirements of the job, however, are consistent for everyone, both in principle and in law under ADA. The capacity to do more than is required for the job is actually irrelevant to work performance and to fair selection under ADA. It would therefore be discriminatory to exclude one applicant from a job assignment just because another applicant could do more, when both could do the job quite adequately.

Functional capacity evaluations have many pitfalls. Because proprietary functional capacity evaluations are comprehensive protocols, not job-specific simulations, they generate a lot of data that are

irrelevant to the employment relationship and which employers collect at considerable risk, because they must then be able to demonstrate that although it was collected, most of the data were not in fact used in making decisions about employment and job assignment. EEOC (which interprets ADA) would almost certainly consider it a violation if an applicant were excluded on the basis of a modality that could not be demonstrated to be directly relevant to a specific job requirement or excluded solely on the basis of poor performance on any modality compared to population norms rather than some validated performance standard.

Some employers have sought to get around ADA by using functional capacity evaluation as a screening tool to demonstrate ability to do the work before a job offer is made. This is possible because under ADA, as interpreted by the EEOC, a test of job-related skill, such as agility, was not a “medical inquiry” or “medical examination.” However, the inclusion of any test that monitored the worker-applicant’s response to the job-related activity, such as blood pressure determination after exertion, or the use of medical equipment made the entire process a “pre-offer” medical examination in violation of ADA. This is a particularly risky skirting of the law if the person conducting or interpreting the test is a health professional.

If a functional capacity evaluation is required as a post-offer evaluation, they may replace or supplement a medical examination, but the same problem of relevance applies. If modalities of the test are not directly related to the requirements of the job and do not reflect a legitimate business interest, then they should not be taken into consideration. This makes most of the functional capacity evaluation a rather expensive irrelevancy.

The physical measurements of a functional capacity evaluation are made in order to ensure a match between the capacity of the worker and the job requirements. Many tests of physical capacity are not necessary for a particular job, and therefore exclusion of an applicant based on inadequate performance on those irrelevant measurements in a functional capacity evaluation would be in violation of ADA. Most jobs in the modern economy do not require such finely tuned

matching in any case. The most practical way to tell if a worker can do the job is simply to have him or her do it.

Another problem is the limited duration of functional capacity evaluations. Many work requirements show training effects: increased skill, improved hand-eye coordination, muscle strength. Evaluating a modality in one sitting cannot capture the training effect. Single-session tests are also irrelevant to the risk of repetitive movement.

Notwithstanding their many limitations, comprehensive functional capacity evaluation may be useful in certain specific applications:

- To evaluate ability and maximal effort in sports medicine, and to document training effects
- To evaluate the capabilities of a patient with severe, multiple, or subtle impairment (for example after recovering from multiple trauma or brain injury) to determine what they can do physically
- To validate a level of impairment where there are many inter-related factors involved
- To resolve a dispute over degree of impairment in which more than one body part and modality are involved
- To evaluate capacity to function under extreme conditions in the military.

When the term “functional capacity testing” is used today by a naïve speaker, he or she often really means impairment assessment.

## **IMPAIRMENT AND DISABILITY ASSESSMENT**

“Impairment assessment” is a medical determination and quantification of the loss of function that limits the activities of a disabled person. Disability is a determination based in part on the impairment assessment that interprets the implications of the impairment for participation in activities of life and work. Disability reflects not only the degree of mental and physical impairment but employability, loss of future income, education, retraining prospects, and the job market.

Impairment is the complement of functional capacity. “Impairment assessment” focuses on identifying and quantifying deficiencies in function as objectively as possible. If functional capacity evaluation, as an ideal, would describe everything that a person can do, impairment assessment as an ideal would describe in precise terms what a person cannot do that they normally should be able to do. Impairment assessment also measures the degree to which a person falls short of a standard, which may be either the population norm, their personal baseline, or what is required for a specific task. The measurement can be translated into a “rating” (in workers’ compensation terminology) that can be used to compare apples and oranges: degrees of impairment of the one part of the body with an unlike other part, such as hands compared to the brain. This rating has applications in disability evaluation and compensation.

Impairment assessment is most heavily used in workers’ compensation and disability programs, and the vocabulary of impairment assessment reflects its origins. However, impairment assessment has more general utility than is commonly recognized. For example, it could be used more than it is now to guide accommodation, to quantify the gap between the level of function that is required and the level of function possible, in personal injury cases to quantify damages, and to study the patterns by level of disability in populations.

When complete objectivity is not achievable, impairment assessment seeks to be at least consistent. Many contributing factors to impairment are not objectively quantifiable, including pain, shortness of breath, and fatigue. Others may be measurable, but the scale may not reflect functionality: for example, the last few degrees of range of motion of the thumb are much more important to the functionality of grip incrementally than the sweep of the range up to the end.

The objective of impairment assessment is to determine what the person cannot do and to measure the deviation from “normal” or intact function or decrements from the pre-existing situation when there was already an impairment in the first place. The impairment evaluation is then used in workers’ compensation of “disability evaluation” to take the next step of matching the impairment against

requirements of the job. This comparison leads to a determination of disability. Disability, however, is a term that applies to the implications of the impairment for activities of living, social functioning, earning an income, and pursuing one's usual occupation. It is a description of the effect of impairment, determined by a complex derivation that takes more factors into account than impairment.

The comparison or standard against which function is compared can be the person's own baseline (which is usually not documented before an injury), normal physiology, norms statistically derived from large populations (in which case the selection of the reference population becomes an issue), or functional requirements of universally required activities of daily living (benchmarked against a threshold for disability and a gradation of increasing significance with progressive loss of function). Ideally, the two most useful standards would be loss of function compared to the individual's own baseline, which would provide an indication of compensable loss, and the threshold for disability, which would establish the significance of impairment. In practice, impairment is usually measured against normal physiology by the use of guidelines, of which the most important are the *AMA Guidelines for the Evaluation of Permanent Impairment* for workers' compensation and *Disability Evaluation Under Social Security* (the "Blue Book") for the Social Security Administration. (There is no counterpart for the Canada Pension Plan.)

## Definitions and Concepts

Definitions are not settled for three essential terms: impairment, disability, and handicap. In general, in North America, "impairment" means a physical or mental limitation on activity, "disability" means the implications of the impairment to the degree that it interferes with social expectations and employment, and "handicap" refers to the gap between what the person can do and what is required.

Different systems within different countries use the same words differently, and it is important to understand what is actually meant, because there is a tension between two opposing points of view. On

the one hand is the view, which underlies ADA, that impairment is a characteristic of the person, subject to mitigation by medical means or rehabilitation, and that the disability is that which remains after treatment and rehabilitation and which defines that person's interaction with the world. On the other hand is the view adopted by most other economically developed countries that impairment and disability primarily reflect barriers to participating in work, social, and family life, and that when a person cannot participate fully because of reduced physical or mental capacity, it is the barriers that should be lowered. This is reflected in a set of subsidiary definitions worked out by the World Health Organization (WHO), which holds that impairments, disability or activity limitations, and barriers to participation are all dimensions of the interaction between the body and the environment, including the social environment: "Impairments are interactions affecting the body; activity limitations are interactions affecting individual's actions or behavior; participation restrictions are interactions affecting person's experience of life."

There are many formal definitions of impairment, the most important of which derive from the disability movement (especially arising out of the United Nations Decade of Disabled Persons, 1983–1992), WHO (especially the International Classification of Functioning, Disabilities and Health, 2001, known as ICF), imputed definitions implicit in ADA (and the earlier Rehabilitation Act of 1973), the Social Security Administration, Canada Pension Plan, state and provincial workers' compensation acts, and in various authoritative guides such as the *AMA Guides to the Evaluation of Permanent Impairment*. The United Nations Decade of Disabled Persons adopted definitions of impairment, disability, and handicap in 1983 that have proven very influential: Impairment is "any loss or abnormality of psychological, or anatomical structure or function. An impairment [is] any loss or deviation of physiological, neurological or anatomical structure or function of an organ or body part (organ and body dimension), a physiological disorder or injury." ADA does not differentiate disability and impairment in functional terms but treats disability as arising from impairment and so implicitly defines it as a

permanent derangement of function. Essentially all defining bodies agree but do not make explicit in their definitions that certain “normal” states that may impose limits for a period are not true impairments, for example, pregnancy, childhood, and advanced age. A synthesis of the various definitions would yield something like the following: Impairment is a loss of capacity, arising from an abnormality in structure or a derangement in physiological function that is clearly identifiable as a deviation either from normal or from that person’s own baseline and which imposes a new limit on activity.

A comparable definition of disability would take into account the implications for the person’s social role. In the United States, disability is generally perceived as a characteristic that attaches to the person and differentiates them from “normal” society. In Canada, Europe, and the United Nations system, disability is a stage or quality of life, representing the interaction between the person and the physical and social environment. The older WHO definition (1980) was “a disability reflect[ing] any limitation or lack of ability that a person experiences in performing an activity in the manner or within the range considered normal for a person, in other words, a limitation in learning, speaking, walking or some other activity (individual dimension).” The current WHO definition of people with disabilities holds that “[p]ersons with disabilities include those who have long-term physical, mental, intellectual, or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others.” The United Nations adopted the Convention on the Rights of Persons with Disabilities in January 2009 and it is currently undergoing ratification by member states. In effect, it has redefined disability as arising from the environment, not from the person’s capacity to function.

The ADA, on the other hand, does not distinguish between impairment and disability, which is a serious conceptual drawback of the legislation. The current (2008 amended) ADA definition of disability is “a physical or mental impairment that substantially limits

one or more major life activities of such individual; a record of such an impairment; or being regarded as having such an impairment.” A comparable synthetic definition, given that the American and the international views are very different, might therefore read something like “disability is reduced capacity, imposed by an impairment, to manage the barriers and challenges within society, the family, employment, or as required for the pursuit of employment prospects, such that the person is impeded in performing normal activities and duties and the normal activities of daily living (taking into account the person’s age).”

“Handicap” is out of fashion as a term, and its meaning has been conflated with disability. This is unfortunate because historically the term means something quite different and useful. Handicap refers to a seventeenth-century game (“hand i’ the cap”) that involved placing bets against unequal odds; the term later was applied to horseracing and golf to refer to adjustment of the gap between unequal players. This is a very useful concept: It recognizes a differential between expectation and capacity to perform that, if remedied, gives all players equal opportunity. Use of the term as a synonym for disability came much later, at the beginning of the twentieth century. (The story that it refers to a beggar holding a “cap in hand” is not true.)

Most formal definitions of handicap, however, equate it with disability and therefore obscure an important distinction between functional impairment and its significance in the real world. The Disability Decade definition from 1983 was “Handicap [is a] disadvantage for a given individual, resulting from an impairment or disability, that limits or prevents the fulfillment of a role that is normal, depending on age, sex, social and cultural factors, for that individual.” The WHO (1980) definition is “Handicap [is] loss or limitation of opportunities to take part in the life of the community on an equal level with others; encounter a person with disability and social, physical environment...an inability to accomplish something one might want to do. The term emphasizes

the focus on shortcomings in the environment and in many tasks and activities, [for example: in] education, occupation, information or communication (social dimension)." It would be very useful to go back to the older definition, perhaps as follows: A handicap is a quantification or description of the gap between real and expected performance or between capacity and requirements to perform activities important to social, family, and employment roles and activities of daily life. A handicap can be overcome by accommodations or adaptations that bridge the gap.

Impairment can be temporary or permanent. In workers' compensation, impairment is not assessed until it is permanent (common term of art are "at permanence" or "maximum medical improvement," which is abbreviated MMI). American usage of "impairment" and "disability" assumes that the condition is permanent, as does ADA, so the terms "temporary impairment" and "temporary disability" are rarely seen. International usage makes no such assumption and considers that anyone may pass through periods of temporary impairment or disability, if only and preferably through advancing age, and that every living person has at a minimum experienced the dependency of infancy. International, and especially European, usage therefore recognizes "temporary disability" as a life stage and uses the term much more often.

It often happens that a person has a pre-existing condition that may be asymptomatic, symptomatic but not to a level of impairment, or stable at some level of impairment. A common example is asthma. When a work-related factor (an injury or exposure) makes the asymptomatic condition temporarily symptomatic or makes the symptomatic condition temporarily worse, this is called "exacerbation." If a worker with pre-existing asthma encounters a trigger at work (such as an allergen or heavy dust), he or she may experience an exacerbation of asthma, with bronchospasm, cough and other evidence of airway inflammation lasting for days or at most a few weeks. If a worker with pre-existing asthma encounters an exposure that decompensates the disorder and leads to a permanent change in level of function, this is called an "aggravation." If a worker with pre-

existing asthma encounters an irritating chemical exposure, the worker may experience a more severe worsening of the condition and symptoms that might lead to a permanent condition with worsened symptoms. This would move the disorder to a new baseline of reduced function. This could be classified as a “second injury.” The difference in impairment and disability between exacerbation and aggravation is that the former is temporary and the latter is permanent, but the medical difference is that exacerbation is a short-term perturbation in a stable condition and aggravation destabilizes the condition and moves it to a new and unfavorable level. A second injury would be a more discreet event that added new impairment to the existing impairment.

In Canada, for a person to qualify for Canada Pension Plan (CPP) disability benefits, the impairment must cause a disability that is “severe” (meaning the person is incapable, in the sense of not fit or able, of pursuing, in the sense of actively working in, any “substantial gainful employment,” not just the former job) and “prolonged” (of indefinite duration and likely to keep the claimant from work for at least twelve months). This is different from most other social insurance systems, in that to qualify the applicant cannot be capable of any work for which he or she is otherwise suited, but the disability does not have to be permanent. Benefits are “all or nothing” (i.e., no partial benefits) and not tied to financial need, because the amount is based on contributions to the system. As a consequence, the process required for CPP is streamlined and focused on reviewing the medical information provided by the physician rather than having the physician conduct an impairment assessment.

### **Impairment Assessment Outside Workers’ Compensation**

The OEM physician is often called upon, because of skills developed in workers’ compensation cases, to evaluate impairment for other benefit or insurance programs, which provide income replacement after disability that does not arise from work. These programs may have

a complicated relationship with workers' compensation in individual cases. Such programs include the following:

- Social Security Administration (SSA) disability insurance
- Canada Pension Plan (CPP) disability benefits
- State temporary disability plans
- Long-term disability insurance

SSA and CPP are social insurance programs designed to prevent catastrophic loss of income. They require demonstration of eligibility, applying a set of criteria that assess whether the applicant is working; if not, whether the applicant has the capacity to perform the activities required to work; if so, at what level; and given that level whether employment is likely to be available. SSA and CPP determinations are highly individualized, taking into account age, education, training prospects, and the local job market for any jobs (not the applicant's previous job). Thus, the outcome may be very different for two applicants with identical impairments. Social insurance programs are predicated on employability and therefore recognize disability as present or absent, not rated by degrees of impairment, as in workers' compensation. Social insurance programs assume that the disability is long term, if not permanent, but require periodic re-evaluation to determine whether there has been improvement and if so whether the recipient's functional status is such that benefits can be terminated.

Some states, such as California, provide temporary disability benefits. They are designed to support the income a person may lose because of either a period of illness or recovery from an injury that is not work related.

Long-term disability insurance (LTD) is commercial insurance against permanent disability that does not arise from work. LTD usually is based on incapacity to work, either at the worker's usual occupation or at any occupation. Policies are increasingly stipulating the latter and must be read carefully to determine eligibility criteria.

These programs exclude disability arising from work-related injury or illness. They are therefore, in a sense, complementary to workers' compensation, and between the two a disabled individual with significant limitations should be covered. The reality, however, is that the systems do not coordinate well. An LTD carrier may determine that an injury was work related after it is denied by workers' compensation, or the reverse. When this occurs, the injured person may face a considerable struggle.

### **Assessing Impairment in Workers' Compensation**

All OEM physicians, however, will at some time evaluate patients in one or more workers' compensation system, so the workers' compensation system will be emphasized in this subsection. Most workers' compensations use or at least acknowledge the authority of the *AMA Guides to the Evaluation of Permanent Impairment* (*AMA Guides*), and so frequent reference will be made to that resource. Some states, notably California, use their own guidelines, but a recent evaluation of the California guides showed that it performed less well than the *AMA Guides* in consistency of use and precision. As the *AMA Guides* improve since the sixth edition, it is likely that states will fall into line in recognizing its authority.

A key threshold in the spectrum of disability and therefore the source of important benchmarks in impairment assessment is the capacity to perform "activities of daily living" (ADLs). ADLs have been defined and delineated mostly through research in nursing and occupational therapy, and there are numerous evaluation scales and instruments. Basic ADLs include bathing, dressing and undressing, eating, transferring from bed to chair and back, continence, using a toilet, and walking. Obviously losing the capacity to do any one of these is profoundly disabling. Instrumental ADLs are a higher level of functioning associated with independent living, including light housekeeping, preparing meals, using a telephone, managing individual transactions involving money and counting change, shopping, and taking medication. A higher level of activities are more directly

related to work capacity and involve the capacity to communicate, mobility in the community, a broader spectrum of physical activity, sensory functions, general use of the hand, and other functions. It is this level that the *AMA Guides* addresses.

Ideally, functional loss would be correlated with an objective, ergonomic standard for the ability to participate in life and work activities. For example, the grip strength required in the thumb might be correlated to the minimum required to take care of oneself and to do different types of work. However, that degree of accuracy is unachievable in practice. It is less important that impairment evaluation be accurate than that it be precise, that is reproducible and consistent. (See Chapter 4 for a discussion of the difference between accuracy and precision.) References such as the *AMA Guides* are arbitrary in the sense of the choices that are made for representing impairment but represent a reasonable consensus by which different cases can be evaluated on the same scale and with consistency.

Impairment assessment for workers' compensation is based on "anatomic loss," which is reflected in structural derangement (often called "deviation" in this context) and "functional loss." Although anatomic loss also involves loss of the function of that body part, the term "functional loss" here means a derangement that affects physiological function rather than the physical integrity of the body. An anatomic loss might be a reduced range of motion for the thumb or abduction at the shoulder and underlies assessment of the musculoskeletal system, which is the most commonly evaluated system because injuries are more numerous than diseases. A functional loss might be reduced pulmonary function or impaired sensation as with a neuropathy. Certain types of functional loss are much more difficult to assess, particularly pain (subjective) and asthma (episodic, with many symptoms, such as cough, that are transient and not easily measured). (The term "function" is used from this point on in its usual meaning.)

The tests that are used to measure impairment are mostly the standard, noninvasive tools of medicine. Some special devices have been invented for the purpose of impairment assessment but have not

achieved widespread use. For example, an “inclinometer,” which acted much like a carpenter’s level to measure angles from horizontal, was recommended by the *AMA Guides* for a time. However, the familiar but less exact goniometer was always more commonly used.

The theory of impairment assessment, as applied in the *AMA Guides*, rests on the notion of the “whole person” (formerly “total person”), defined as having no impairment (no structural or functional derangement) and no limitations on activities of daily living or employment. This ideal person is given a rating of 0 percent impairment of the whole person. “Total impairment” refers to an individual with 100 percent impairment of the whole person who is not dead but is impaired to the point of being unable to perform basic activities required of work and daily living. Total impairment may arise from one major impairment, such as cognitive dysfunction due to brain damage, or to the aggregate effects of many partial impairments. Between the two extremes of 0 and 100 percent impairment of the whole person is a spectrum of “partial impairment.”

Partial impairment is rated on the basis of percentage points reflecting loss of function. Most “organ system” or systemic categories of impairment are rated directly as percentage impairment of the whole person, often by classes of impairment. The system for respiratory disorders, for example, reflects a range within classes of impairment based on criteria for pulmonary function tests, developed by the American Thoracic Society. The system for cardiovascular disease includes a range of impairment within the functional classes defined by the American Heart Association.

The system is slightly different for the musculoskeletal system, which is anatomically based. Each body part is assigned a nominal percentage of the whole person. For example, an upper extremity is deemed to be 60 percent of the whole person, which is not unreasonable when one considers how difficult it is to participate in the activities of life and work with only one arm but at the same time how one might be able to adjust. The loss of function is then assigned a percentage specific to that body part. For example, if the upper extremity could not be used through atrophy or ankylosis, the loss of

function for the upper extremity would be 100 percent. Another impairment of the shoulder might result in 50 percent impairment of the upper extremity. An impairment of the hand and wrist might result in 30 percent impairment. These impairments of the body part are then applied to the percentage of the whole person in order to derive the percentages of the whole body. In the example given of the upper extremity, which is 60 percent of the whole person, a 50 percent impairment of the upper extremity would yield 30 percent impairment of the whole person, and 30 percent impairment of the upper extremity would yield 18 percent impairment of the whole person.

There are often multiple impairments, either of different measurements on the same body part (such as strength and range of motion) or different body parts (such as the upper extremity and the lower extremity). Because 100 percent is an upper bound of impairment of the whole person and because multiple impairments tend to reduce to a more general pattern of impairment, they cannot be simply added. The *AMA Guides* uses a “combining table” that combines the values without adding them arithmetically. For example, the 18 percent impairment of the upper extremity combined with a 30 percent disability would yield a disability of 43 percent, still high but not as high as 48 percent.

“Second injuries” are subsequent injuries after the first to the same body part that lead to cumulative impairment. The principle is to measure the impairment after the second injury is at permanence and then subtract the impairment that was documented from the first injury. The difference is the impairment attributable to the second injury. Second injuries have important implications for compensation management.

The individual protocols place appropriately greater weight on functional limitation than on anatomic derangement. For example, the impairment ratings for the thumb are not linear over adduction, that is, across the entire sweep of the thumb over the palm. The rating goes up exponentially as the thumb approaches neutral position. This is because the thumb is needed to grip, and this function is much more important than full adduction.

The *AMA Guides* place great weight on activities and functions that are important in activities of life and work. They address that part of the spectrum of impairment that reflects the capacity to do work and are therefore most useful at lower levels of impairment. Other instruments would be needed to evaluate degrees of severe disability, for example for nursing or assisted-care services.

The *AMA Guides* have gone through many editions. Many states require the most current edition (currently the sixth), but some states require the fourth because the fifth was widely perceived as a failed product and inter-rater reliability was poor. The sixth edition has moved strongly in the direction of compatibility with the ICF and incorporation of modern insights into functional capacity and is also easier to use. The *AMA Guides* are still complicated and require training to use.

Two organizations provide comprehensive training in impairment assessment and use of the *AMA Guides*. The American Academy of Disability Evaluating Physicians (AADEP) is a membership organization, consisting of a variety of specialists, that offers comprehensive training in the form of continuing education and certifies practitioners by examination (Certification of Evaluation in Disability and Impairment Rating, CEDIR). AADEP is particularly strong in Texas, where the state workers' compensation system uses a "designated doctor" system of referrals for impairment assessment and has relied on AADEP training. The American Board of Independent Medical Examiners (ABIME) is organized as an autonomous body providing training courses in all aspects of independent medical evaluation but emphasizing impairment assessment. It certifies practitioners by examination for its own credential (Certified Independent Medical Examiner, CIME). (ABIME is not a medical specialty board recognized by the American Board of Medical Specialties.) ABIME may be more widely recognized, has made inroads in some other countries, and is more prominent in occupational medicine. AADEP and ABIME have certified comparable numbers of physicians and are equally respected. In Canada, the Canadian Society of Medical Evaluators is a membership organization that certifies on the basis of submitted examples of impairment assessments.

## ACCOMMODATION

Returning to the early, still useful definition of handicap, bridging the gap between what a disabled person can do and what is required can be viewed either as equipping the worker to make up for a limitation in function or lowering the barriers to allow participation in an activity. The means to do this is “accommodation,” which is simply working out a plan or modification that allows a worker with a disability to do the work.

As the working population ages, accommodation may become less a specific intervention for a person with disabilities than a factor in workplace design for safety and productivity for all workers. Accommodation, by definition, is specific to the individual. Every disabled person’s disability is unique to their own combination of work capacity, skills, impairment, and personal adaptations to the impairment. However, many disabled persons have similar needs, and tools, workplaces, and equipment can be designed in the first instance to accommodate workers with or without disabilities. This concept of “universal design” has become increasingly influential in design and engineering as it has become apparent that many accommodations for disabled persons are also easier and more efficient for persons without disabilities. Good ergonomic design, which is at the root of universal design, also reduces injury risk. For example, scissors are particularly difficult for people with hand disabilities or arthritis to use. Scissors designed with flared handles and with the blades at an angle to the handle instead of straight are much easier to use for workers who are disabled but also for those who are not, and these designs greatly reduce fatigue with repeated use. They are also likely to reduce the risk of repetitive strain injury. Universal design may make accommodation unnecessary for many, perhaps most, situations in which it would otherwise be required while at the same time allowing greater productivity and reducing barriers to employment as workers age.

“Accommodation” is a term that predated ADA, but its meaning as a means to bridge the gap between disability and job requirements

has been redefined and fixed by ADA, which requires “reasonable accommodation” to a qualified person with a disability in order to meet the essential requirements of the job. Accommodation is usually considered in the context of new hires and qualified applicants for the job assignment. Accommodation may also be required for return to work (RTW), when a recovering injured worker returns to modified duty, either in job-related activities or shorter working hours. In this context, most modified duty for RTW is temporary and bridges the gap created by temporary impairment. Sometimes, however, a worker who has a new permanent impairment will require a permanent accommodation.

An accommodation may be any of the following:

- An assistive device, such as large screen for the vision impaired, an amplification system for the hearing impaired, or ergonomic tools that are easier to grasp and use (particularly scissors)
- Companion animals, specifically guide dogs for the blind, which would be considered an accommodation modifying a policy against pets at work because guide dogs are not kept as pets
- Workplace modification, such as a wheelchair-friendly workstation or enhanced illumination
- Rebalancing job assignments so that the person with the disability can do those parts within their capacity and other workers can do the rest
- Restructuring work organization so that medical and rehabilitation appointments can be kept without disruption, responsive leave policies, and modified work hours, if needed
- Modification of equipment, such as ergonomic adaptations and interventions for persons who do not have full use of their hands
- Removing barriers to access to the workplace, such as reserved parking spaces for the disabled or use of a different entry point
- Removing barriers to access at the workplace, such as placing files in lower drawers for easier access to persons in wheelchairs,

providing Braille labels beside elevator buttons, or ensuring that there are no barriers to washrooms

- Adjustment of examinations and evaluations, as for a worker with a cognitive impairment who may require more time to finish a test (but a timed test might be acceptable if it addresses a specific work requirement, for example, the ability to read and act on messages rapidly, and is given to all applicants)
- Training materials and policies for co-workers to help them understand the implications of disability
- A dedicated assistant, such as an amanuensis to take dictation for a worker who cannot physically write, a sign-language interpreter for a worker who is deaf, a page-turner for someone who cannot use their hands, a travel assistant for required business trips, or a reader for a worker who is blind
- Communications devices, such as Text Telephone (TTY) and Braille devices

“Reassignment to a vacant position” is an accommodation mentioned in ADA that would apply to current employees. Reassignment to another, vacant position should not be considered unless it has not been possible to make a reasonable accommodation to the employee. The new position should be equivalent in terms of responsibility, pay, and status if at all possible. An employer is allowed to reassign a worker with a disability to a lower-status position if accommodation would impose an undue hardship, but this is likely to cause problems and to raise suspicion that it is being done to encourage the worker to leave (as a form of “constructive dismissal”). Legal guidance from an expert on ADA and employment law is advised for any employer considering this option.

Facilities for employees, such as health centers, cafeterias, lounges, and fitness centers also fall under ADA, and similar accommodations are required to allow persons with disability to use the facilities.

Employers are not required under ADA to supply medical or prosthetic devices that are required for activities of daily living. For

example, employers would not be expected to provide eyeglasses for a person with vision impairment, unless they were somehow specific to requirements of the work. Employers are not expected to provide amenities (such as air fresheners), conveniences (such as an office refrigerator), or personal products (such as tissues) that are not provided to employees without disabilities.

ADA requires reasonable accommodation, defining “reasonable” as that which does not cause undue hardship. “Undue hardship” might include excessive cost, an investment disproportionate to the size of the facility and the number of people working there, or one causing detrimental impact on operations at the facility. An employer should consider alternative reasonable accommodations before concluding that an accommodation would impose an undue hardship; an accommodation does not have to be ideal.

The best accommodations are simple and sustainable and are designed in partnership with the employee, who is the best expert on his or her specific disability. Most accommodations under ADA cost little or nothing. It costs little to reassign work. Whenever possible, commercial “off-the-shelf solutions” are preferred because they are less expensive than custom-built devices and are supported by the manufacturer. Tailored, jerry-rigged, or custom solutions are more likely to be expensive, take more effort to design, are less likely to be successful, and quite possibly may cause hard feelings if they fail after the effort has been expended in good faith.

Most employers do not object to minor modifications in the workplace, but they often do not know how to make these modifications. Fortunately employers in the United States and Canada have access to assistance in the form of the Job Accommodation Network (JAN), a project of the President’s Committee on Employment of the Handicapped. JAN is coordinated out of an office on the campus of West Virginia University that serves as a clearinghouse for ideas and suggestions regarding work modifications to accommodate the handicapped or the temporarily impaired. JAN has an extensive experience with workplace modifications, some of which are very simple to make and at low cost. They can be reached by

calling 1-800-526-7234 (or 1-800-526-4698 inside West Virginia), a toll-free number. Their services are free of charge. Examples of the modifications they may recommend include use of a slightly different tool, a change in the seating arrangements, providing a different type of switching mechanism, a special chair, and installing lever handles rather than door knobs on doors for individuals who have trouble with grasp.

Accommodation requires recognition of the need by both employer and employee and cooperation between them to reach a practical, sustainable solution. Most employers are understandably reticent to raise the issue at the time of the job interview because it may be construed as a pre-employment medical inquiry, which is not allowed. Having the physician evaluate the need for accommodation as part of the pre-placement evaluation is the solution for applicants. For current workers, ADA allows employers to inquire about accommodations that may be needed if a problem is obvious and business related. Most large employers have policies that put the responsibility on the worker to request accommodation if it is needed. Sometimes workers may be reluctant to raise the issue with their immediate supervisor. The policy should state that the worker has the option of discussing accommodation with a representative of the employer's department of human resources.

## **DRUG SCREENING**

Drug screening has become a major emphasis in occupational medicine in the United States. Not every OEM physician will be involved in drug screening, but all should understand it in principle because it is a common and critical service, particularly in certain important industries such as transportation.

Drug use is considered to be a voluntary form of work incapacity, one that has the potential to place the worker and others at risk, particularly when the worker is assigned to a "safety-sensitive" job, one in which incapacity may present a direct threat to the public or other workers. In the 1980s, there was also concern that high levels

of drug use in the community were spilling over into the workplace, causing a serious risk to the public, encouraging drug abuse, and reducing national productivity. As a result, many employers introduced drug screening programs, and federal agencies began to require drug screening for safety-sensitive positions under federal jurisdiction. At first, there was great opposition to the institution of drug testing programs, on grounds of human rights, administrative complexity, legal liability, and ethical pitfalls. However, the initial high rates of screening appeared to validate the view that screening was necessary. Although detection rates are now much lower, drug screening is a permanent fixture of working life in many industries, particularly transportation, and has become driven by federal legislation, although many employers continue to be motivated by social attitudes regarding drug use among their employees.

Drug abuse follows regional trends. Drug use in the workplace was thought to be mostly an urban phenomenon when widespread drug screening began. Alcohol abuse is ubiquitous but is a particular problem in isolated and remote workplaces and in rural and far northern regions. For many years amphetamine usage occurred predominantly on the U.S. West Coast. However, the current methamphetamine epidemic in the American middle states and the rapid spread of heroin in rural areas has made the problem more geographically dispersed in North America. Drug abuse may have different implications in rural settings, for example, because of longer driving times, which increases the risk of motor vehicle incidents. Drug abuse also both reflects and contributes to local patterns of crime, domestic violence, mental health services, and outlets for more constructive community activity, especially for young people.

Drug screening is uncommon in Canada, except among subsidiaries of American companies, among oil companies, and by transportation companies that carry people and goods across the border into the United States, where they are subject to U.S. Department of Transportation regulations. In 1994, the Government of Canada decided not to pursue legislation empowering Canada Transport to require drug screening, as in the United States. Individual companies

were free to establish their own programs within a set of guidelines established in 1988 by the Canadian Human Rights Commission, which requires that the testing be relevant to the essential requirements of the job (the term of art is “*bona fide occupational requirement*,” BFOR), that there be a demonstrable need to ensure a drug-free workplace (safety being the obvious need), that testing meets that need, and that there be a system for referral and accommodation, such as an employee assistance program, for workers who test positive. There are also broad requirements for the protection of human rights and privacy. Random testing has been disallowed under case law, but testing on the basis of “reasonable grounds” for suspicion of impairment and pre-placement or pre-employment testing (ADA terminology does not necessarily apply in Canada, of course) have been upheld, with some exceptions. Few employers in the transportation sector have initiated drug screening programs, relying on the threat to drivers of criminal sanctions for operating vehicles while impaired. Employers that have drug screening programs incorporate alcohol by breath testing as well as drug screening in urine.

### **Medical Review Officer**

The primary role of the physician in drug testing is documentation, so that the essential facts are recorded and the decision that is made will withstand scrutiny and future litigation. The U.S. Department of Transportation (DOT), by requiring urine drug testing for transportation workers, particularly commercial drivers, has created a demand for physicians with certified expertise in drug testing and interpretation. Positive screening tests must be confirmed and interpreted by a physician, the “medical review officer” (MRO); issues arising include cross-reactivity or confusion with prescription medication. Because of the potential for serious repercussions in the event of a false positive test, the system relies heavily on quality assurance mechanisms and certification of laboratories by the National Institute of Drug Abuse (NIDA). The system requires the MRO to contact the test-positive subject personally before reporting the result.

The knowledge and professional demeanor of the physician serving as MRO are critical, but the field is changing so rapidly that it is difficult to keep up as an individual without formal continuing education.

The preferred way for OEM physicians to prepare for service as an MRO is to take a two-day course offered twice yearly by the American College of Occupational and Environmental Medicine: the “Medical Review Officer Training Course.” A formal certification examination is offered by the Medical Review Officer Certification Council, and independent, nonprofit certification body. The other major certification body is the American Academy of Medical Review Officers.

Although many occupational physicians perform MRO services as private consultants as part of their overall mix of services, it takes a great deal of organization and a high volume to base an entire practice on this service. The key is a committed system organized with employees who are extremely reliable in handling the paperwork. There have been several proposals to remove the requirement than an MRO be a physician, but DOT has kept this provision because of the need to have some professional accountability guaranteed and for a licensed practitioner to take responsibility in each case for medicolegal purposes.

In essence, drug testing as it is usually practiced is a structured forensic program for the detection of illegal drug use among current or potential employees. The tests are usually performed on urine for the detection of the five drugs identified by NIDA as the priority drugs of abuse among adults in the United States: opiates, cocaine, marijuana, amphetamines, and phencyclidine. Alcohol is conspicuous by its absence, but its use and even abuse in private is not against the law. When alcohol testing is performed at the discretion of the employer, it is by breath test with confirmation of positive results by blood test. The scheduling of such testing is typically on presentation for applying for a job and randomly thereafter among employees after hire, although few employers do routine alcohol testing. Semisynthetic opiates, such as oxycodone, are also absent, although they are common drugs of abuse. Discretionary testing by employers for these drugs does not seem to occur.

Most drug testing programs are set up in compliance with the mandatory programs required by the DOT and rely heavily on NIDA recommendations. In addition, employers may add their own discretionary programs, but these must be kept strictly separate from the federal programs, even to the point of using different urine samples. Drug testing must be handled carefully to prevent legal action for discrimination. Policies should be explicitly documented, and drug testing schedules (random or otherwise) should be equitable for all employees, not applied selectively or arbitrarily.

The initial emphasis in drug testing programs was punitive; the drug-using employee was an “offender” who got “caught.” After several years of experience, the prevailing attitude has become more sympathetic to the employee, with an emphasis on early detection of a drug habit leading to referral to an employee assistance program (EAP, discussed in Chapter 19) and mandatory treatment. Some experts advise keeping mandatory treatment programs for drug test-positive employees separate (with separate records) from EAP services for employees in general, so as not to compromise or create a stigma for the EAP program.

Compliance with drug testing is made a condition of employment by the employer; refusal to test is insubordination and grounds for dismissal. Most employers provide EAP services and mandatory treatment for their drug test-positive employees once, and some provide a second chance; none tolerate repeated positive tests or accept positive tests among job applicants. Evidence that an employee has willfully tampered with a drug test is generally grounds for immediate dismissal.

The profile of the drug-using employee is not always predictable. Some high-producing employees have been demonstrated to get high on cocaine or methamphetamine, and a few managers have even managed impressive careers while on heroin.

The experience of drug testing programs has been changing. Rates of positive testing ran as high as 25 percent in some areas when testing was first introduced in the mid-1980s (which may also have reflected where they were first introduced). Rates now are typically

around 1 percent (as they have been when attempted in Canada and in the U.S. military) or less. Conventional wisdom in the field is that “only stupid employees get caught,” which is harsh but probably not far off.

Most employees who test positive eventually move on to other employers. It is said among some personnel managers that one effect of the program has also been to displace applications from drug-using job seekers away from employers that test and toward nearby employers that do not test. It is said that some employers that do not test have seen much more trouble with drug use among their new hires after major employers in the same community have instituted testing. As well, employers that require testing of employees of their subcontractors often find much higher rates of positive tests than among their own employees. Employers who have instituted drug testing often find that after the initial impact there are diminishing returns to committing further resources or intensifying the testing program. The deterrent effect of random testing seems to be early and all at once. The deterrent effect of pre-placement and nonrandom or openly announced drug testing is minimal, and it has been said, not altogether without cause, that these programs serve mainly to weed out the young, the confused, and the uncreative from the employer’s workforce applicants.

Given the initial emphasis on the magnitude of the problem of drug use to the American economy, it is surprising that the institution of drug testing has had little or no effect on injury rates, health care costs, productivity, or other health or management outcomes.

Some OEM physicians find MRO work distasteful because it is intrusive into the private lives of workers. Others accept it as necessary in the public interest.

### **Technical Aspects of Drug Testing**

Drug testing programs tend to parallel DOT regulations and the recommendations of NIDA, in part to provide a defensible position in the event of litigation.

The technical aspects of drug testing are complicated. Besides the laboratory procedures, which will not be dealt with in detail here, the procedures for obtaining the urine specimen (or "sample") and certifying that it has been guarded to prevent tampering are involved. One should always assume that the findings on the sample will be evidence in court.

Urine collection must be done in a highly structured manner, discouraging efforts to adulterate or dilute the sample. Collections are not usually observed directly unless the employee is being retested after a questionable initial sample that showed evidence of tampering; when this is necessary, it is usually done with mirrors in the toilet stall (and is no less distasteful to all concerned). A large underground business exists in drug-free urine and devices to deliver it into specimen jars while appearing to urinate. Since one effective means of lowering the probability of a positive test result by a subject who is borderline, having used drugs several days before, is to dilute the sample, it is important to prevent access to any source of water. Taping of the toilet tank and coloring of water in the toilet bowl by blue dye is standard practice. The temperature of the urine must be checked and must be within reasonable range of body temperature. The specific gravity is also checked for physiological plausibility.

The urine sample is split, and the second sample retained for retesting in the event of a legal need to do so. The sample, once obtained, must be kept in a strict "chain of custody" with supervision at every step and every transfer documented. In practice, this is accomplished by sealing the sample container, labeling it with a number (not with the employee's name), and following its path on a form in which anyone who handles it affixes his or her signature at the point of transfer. The actual procedure is spelled out in the paperwork at every step.

Testing of the sample is subject to the same factors of clinical epidemiology described in Chapter 5; modern drug testing techniques have high sensitivity and specificity and are usually performed in tandem. A positive radioimmunoassay, for example, will be followed by a combined analysis by gas chromatography and mass spectrometry

(GC/MS). This is expensive but necessary to provide accurate identification for legal purposes and to rule out cross-reactivity. For example, the Vicks® brand nasal inhaler contains the l-isomer of methamphetamine, which is not a psychoactive drug of abuse. Differentiating this from the illicit d-isomer required GC/MS until the introduction of monoclonal radioimmunoassay kits. Over-the-counter nasal decongestants that include ephedrine have also produced confusing cross-reactivity.

The technology of drug testing is increasingly sophisticated. A quantitative analysis may yield evidence of the presence of a drug; above a certain cutoff the test is considered positive. The cutoff is essentially zero detectable for most, but is usually set at 100 mg/l for cannabinoids (marijuana residues). Except for opiates, for which the quantity found is important in assessing the possibility of prescription medicine usage, laboratories report the test as positive or negative for the five drugs of concern.

Samples that are stored, even under refrigeration, may undergo degradation, and retesting may not replicate the initial concentration of drug residues quantitatively. This is a particular problem for marijuana and the cannabinoids.

Interpretation of positive drug tests requires a grounding in toxicokinetics and is beyond the scope of this section. However, a few points will illustrate the complexity of interpretation. Urine testing by conventional means does not detect the common drug residues after more than a few days, even for cannabinoids, which tend to be more persistent. Opiates may be interpreted by reference to the quantity found of codeine compared to morphine; if the ratio is more than an order of magnitude with a predominance of codeine, codeine administration is a plausible explanation for the positive test (except that very late in excretion and at low concentrations the ratio can equalize). Poppy seeds, baked on bread, rolls, or bagels or mixed in muffins are rarely responsible for positive tests; the subject may have been unaware of eating them. However, most positive tests are clearly positive and not so easily explained. Drug-using subjects quite commonly invoke such stories and explanations to cover their drug use.

Urine testing is accompanied at some facilities by plant searches and inspections of lockers, vehicles, and the personal property of workers. The legal implications of such searches should be reviewed with legal counsel before considering such programs.

### **Department of Transportation**

The Department of Transportation (DOT) mandates drug testing for four million workers in the transportation industry covered by the Federal Highway Administration (mostly interstate truckers), the Federal Aviation Administration, and the Coast Guard, as well as workers on interstate pipelines. There is also an internal DOT testing program for the agency's own employees, mostly air traffic controllers; it is used by the agency as a model for development of national regulations.

Testing is required following accidents, with reasonable cause (to suspect the influence), pre-employment, and at random, and may also be performed on a periodic basis when these workers receive their medical evaluations. The DOT system emphasizes confidentiality (to the point of prohibiting identification of the employee to the laboratory except by number) and is exclusively for DOT purposes; no other tests can be performed on the sample. Information on medications being taken are obtained by the MRO only after a positive test is obtained and cannot be collected in advance as part of the drug test. Testing must be done at a NIDA-certified laboratory. On-site drug testing is not yet permitted.

All drug testing reports must be received by the MRO in hard copy, not orally in person nor by telephone; employers cannot receive reports directly. (Self-employed individuals may receive information about themselves.) Positive test results must be signed by the certifying laboratory scientist; negative tests do not need to be certified. Positive screening tests are confirmed by GC/MS. Medical interviews and interpretation to investigate positive tests can only be performed by licensed medical practitioners acting in the capacity of MROs. This investigation must take place before the test result is

reported to the employer and may require tracking down the employee while maintaining confidentiality as to why. The employee, who frequently challenges the test result, may demand a retest of the original sample within seventy-two hours for (nonquantitative) confirmation. Employers are entitled to know which drug tested positive, but not how much was found.

DOT regulations require the employer to remove the test-positive employee from the critical safety function in which that person is serving, effectively suspending him or her from driving, flying, and piloting. Positive tests among aircraft pilots must also be reported to the Federal Aviation Administration, and a determination must be made whether or not the pilot is dependent on drugs. DOT regulations do not require referral to EAPs or mandatory treatment.

The paperwork required for DOT drug tests is extensive and legally precise. There are several tricky aspects to the current paperwork that could invalidate the test documentation if performed incorrectly.

### **Other Government Agencies**

Mandatory drug testing is also required by the U.S. Nuclear Regulatory Commission for operators of nuclear power stations and other personnel handling critical radiation hazards. The paperwork protocol for testing differs from that of DOT. The Department of Defense has a drug-testing program for military recruits and may have programs for other personnel.

Some states have imposed drug testing regulations on intrastate drivers similar to the DOT regulations for interstate drivers. State legislation may change, and inquiries should be made locally before making assumptions.

In the occupational setting, prevention takes the form of worksite programs. Typically, occupational approaches to prevention are blended, as in matching noise control (hazard control) and hearing conservation programs (surveillance). New and more effective strategies for delivering preventive services have been developed in recent years that lend

themselves to introduction in the workplace. However, the implementation of these strategies is often constrained by business considerations, distrust of management, questions about cost effectiveness, and the contradictory health messages and choices that are embedded in society as a whole.

## **RESOURCES**

- American Medical Association. *AMA Guides to the Evaluation of Permanent Impairment*. 6th ed. Chicago: American Medical Association; 2008.
- Hartenbaum N, ed. *The DOT Medical Examination*. 3rd ed. Beverley Farms, MA: OEM Press; 2003.
- Roelen CAM, Koopmans PC, Notenbomer A, Groothoff JW. Job satisfaction and sickness absence: a questionnaire survey. *Occup Med*. 2008;58(8):567–571.
- Talmage JB, Melhorn JH. *A Physician's Guide to Return to Work*. Chicago: American Medical Association; 2005.
- Taylor NAS, Groeller H. *Physiological Bases of Human Performance during Work and Exercise*. Edinburgh: Churchill Livingston, Elsevier; 2008.
- Thomas VL, Gostin LO. The Americans with Disabilities Act: shattered aspirations and new hope. *JAMA*. 2009;301:95–97.

## **NOTEWORTHY READINGS**

- Anonymous. Workers' health—occupational or personal? *Lancet*. 1984;i:1390–1391.
- Baker CC, Pocock SJ. Ethnic differences in certified sickness absence. *Brit J Indust Med*. 1982;39:277–282.
- Carey TS, Hadler NM. The role of the primary physician in disability determination for social security insurance and workers' compensation. *Ann Intern Med*. 1986;104:706–710.
- Cowell JWF. Guidelines for fitness-for-duty examinations. *Can Med Assn J*. 1986;135:983–988.
- Cox RAF, Edwards FC, Palmer K. *Fitness for Work: The Medical Aspects*. Oxford: Oxford University Press; 2000.
- DeHart RL, ed. Medication-induced performance decrements. *J Occup Med*. 1990;32(4):special issue.
- EU-MAHDIE Working Group. Measuring Health and Disability in Europe: Supporting Policy Development. A European Coordination Action for Policy Support. 2006. Available at [www.mhadie.it/getDocument.aspx?FileID=158](http://www.mhadie.it/getDocument.aspx?FileID=158).
- Guidotti TL, Cowell JWF, Jamieson GG. *Occupational Health Services: A Practical Approach*. Chicago: American Medical Association; 1989.

- Hartigan JA, Wigdor AK, eds. *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery*. Washington, DC: National Academy Press; 1989.
- Himmelstein JS, Pransky GS, eds. Worker fitness and risk evaluations. *Occup Med: State of Art Rev*. 1988;3(2):169–364.
- Holleman WL, Holleman MC. School and work release evaluations. *JAMA*. 1988;260(24):3629–3634.
- de Kort WLAM, Post Uiterweer HW, van Dijk JH. Agreement on medical fitness for a job. *Scand J Work Environ Health*. 1992;18:246–251.
- Leonardi M, Bickenbach J, Ustun TB, Kostanjsek N, Chatterji S, MHADIE Consortium. The definition of disability: what's in a name? *Lancet*. Oct 7, 2006;368:1219–1221.
- McKeown KD, Furness JA. Sickness absence patterns of 5000 NHS staff employed within Northallerton and South West Durham Health Authorities. *J Soc Occup Med*. 1987;37:111–116.
- Pransky GS, Dempsey PG. Practical aspects of functional capacity evaluations. *J Occup Rehab* 2004;14:217–229.
- Reville RT, Seabury SA, Neuhauser FW, Burton JF Jr, Greenberg MD. *An Evaluation of California's Permanent Disability Rating System*. Santa Monica CA: RAND Institute for Civil Justice; 2005.
- Shepherd RJ. Assessment of occupational fitness in the context of human rights legislation. *Can J Sports Sci*. 1990;15:89–95.
- Social Security Administration. Disability Evaluation Under Social Security. SSA Pub No 64-039. September 2008. Available at <http://www.ssa.gov/disability/professionals/bluebook>.
- Task Force on Health Surveillance of Workers. Health surveillance of workers. *Can J Pub Health*. 1986;77:91–99.
- Taylor PJ. Occupational and regional associations of death, disablement, and sickness absence among Post Office staff 1972–75. *Brit J Indust Med*. 1976;33: 230–235.
- Taylor PJ, Burridge J. Trends in death, disablement, and sickness absence in the British Post Office since 1891. *Brit J Indust Med*. 1982;39:1–10.
- Taylor PJ, Pocock SJ. Sickness absence—its measurement and control. In: Schilling RSF. *Occupational Health Practice*. London: Butterworths; 1981: 339–358.

*This page intentionally left blank*

# **19 WELLNESS AND HEALTH PROMOTION**

Occupational and environmental medicine (OEM) is a specialty of preventive medicine. A fundamental concern of occupational and environmental medicine is to protect the health of workers by preventive action, most obviously by preventing or controlling exposure to hazards. This chapter deals with interventions that are intended, in general, to change individual health risk, mostly by changing personal behavior and lifestyle. These organized programs are variously called wellness, health promotion, and applied prevention. They apply the theories and insights of prevention science (see Chapter 8). The ideal is, whenever possible, to go further than disease prevention and to strive for achieving gains in individual health and fitness. Applied to disease management, this includes improving the health of people with chronic diseases (such as diabetes) and preventing complications.

On the occupational medicine side, much effort is devoted to organizing and administering programs intended to keep workers healthy, especially in large corporations. In recent years, wellness programs have become increasingly prominent in occupational medicine practice and have become an integral part of the OEM physician's responsibility. The essential idea behind wellness and

health promotion programs is that there are health gains to be achieved by applying the resources, technology, and management available through large organizations to the objective of helping workers or community residents, as individuals or as groups, to enhance their health. This is also, of course, the fundamental idea behind managed care in another context.

On the environmental medicine side, wellness and health promotion (at least as expounded by the Ottawa Charter, as described below) are largely a matter of identifying and protecting the community against exposure to environmental hazards. However, community-based programs can also enhance health by promoting environmental interventions that are under the control of individuals, families, and communities, such as injury prevention or preventing childhood lead exposure. They overlap with other community-based health promotion, screening, general prevention, and public health programs. This topic will be addressed only to a limited extent in this set of volumes. The major emphasis will be on the traditional role of prevention in modifying individual health risk factors and of health promotion in seeking health gains.

In occupational and environmental medicine, physicians are called upon to advise on health protection, to partner with other health professionals in the prevention of exposure to hazards, to conduct periodic health evaluations and surveillance programs for exposure-related disorders, to support health promotion or wellness programs, and to advocate control of exposures to well below permissible exposure levels, an example of the “public health” approach.

## **HISTORICAL CONTEXT**

Preventive medicine is as old as medicine, and its history, and that of environmental health protection, is beyond the scope of this chapter. The idea that the health of workers could be protected and improved by interventions and services delivered through the workplace is newer and has its roots in the social disruption that accompanied the work-related hazards introduced by technology and rapid

economic growth in Europe and North America in the nineteenth century.

The idea of protecting the general (as opposed to occupational, or work-related) health of workers and their families has a long history going back to utopian and industrial reformers of the nineteenth century. During the Industrial Revolution, living conditions fell dramatically for the new working class in many (but not all) new industrial cities, and large numbers of immigrants strained charity and social services. The predominant social philosophy of the day was *laissez-faire* economics and “survival of the fittest” (sometimes called “social Darwinism” but in fact predating Darwin and the probable source of inspiration for his theory of natural selection). By this philosophy, workers had to expect to endure hardship. If the weak did not survive, the process of selection supposedly made society all the stronger. As a general rule, efforts to reform conditions of the working class took an ideological turn in Europe but kept to a less ideological, reformist path in the United Kingdom and the United States, although there were many exceptions. In Europe, the reaction against the threat of revolution in 1848 ultimately included many concessions and reforms, especially around 1880, when the ideological momentum for a political solution to workers’ conditions was waning. By the end of the nineteenth century, it was revived again in Europe by Marxism.

Although conditions of the working class were a political issue during the nineteenth century, some employers, chief among them Robert Owen (1771–1858), who was widely influential in both Britain and the United States, genuinely felt a strong obligation to improve the lives of their employees and introduced reforms in working conditions, nutrition, and programs to discourage alcohol abuse. He established utopian communities for his workers that ultimately were unsustainable and failed. Owen later refined his political beliefs into a program of social action that is considered by many to be the forerunner of modern socialism, notwithstanding that he was a prominent and successful capitalist. Other employers, such as George Pullman, used their power as employers to coerce workers into what they considered to be healthy and responsible lifestyles.

Pullman, who invented and manufactured luxury sleeper cars on trains, founded a utopian community for his employees that became Pullman, Washington, in the 1880s, which later experienced a historic workers' strike against autocratic rules governing health, hygiene, and civil liberties. Underlying these ideas was the notion that working people could not take care of themselves and needed guidance from a paternalistic employer.

By today's standards, these measures were uncomfortably patronizing in tone, but they often, as in the case of Owen's reforms, represented dramatic improvements in the quality of life of working people compared to their less fortunate neighbors who worked for employers with then-conventional views. On the other hand, these experiences created distrust in the vision and motivation of employers. This distrust is still reflected in the skeptical attitude of organized labor toward health promotion programs.

One humanitarian response to deplorable working conditions in the nineteenth century was the "settlement movement" in the United States, which was particularly strong in Chicago. The settlement movement was led by Jane Addams, a formidable but class-conscious reformer who was concerned with raising living standards for the working poor. Addams was mentor to Alice Hamilton, the great occupational physician and toxicologist (see Chapter 1). Her goal with respect to health education and protection was to help new arrivals settle their families in decent living quarters, teach them to eat a balanced diet, and provide medical care for their children. The settlement movement undoubtedly did a lot of good, but what really changed the picture of working life in North America on a large scale was education and rising wages.

Modern concepts of workplace-centered health and wellness were developed by Harry Mock around 1907. Mock was a luminary in occupational medicine, also working in Chicago (see Chapter 1). In a seminal textbook, he argued that workers and employers benefit from health protection and that reducing health risks brought a more stable and productive civil society. He initiated educational programs, direct assistance to workers and their families, and programs to address the

needs of vulnerable populations (specifically, working women and immigrant groups) at Sears Roebuck & Co., where he was medical director. The assumption was that a large company, with all its resources, efficiencies, and capital, could achieve health gains for its workers through organization, training, and economies of scale. The roots of managed care can be found in this concept, no less than the forms of prepaid care and closed healthcare delivery systems that had already been introduced by railroads and some other large employers.

Concepts of population health, although the term was not used then, came to public attention during both world wars when, at the beginning of each war, U.S. Army recruits were found to be in generally poor health. If presumably able-bodied young men were in such bad shape, the reasoning went, the health of the nation as a whole must have been dismal. This realization that the health of the population was linked to readiness caused concern and led to studies and reports, but options for action were limited during wartime. They became even more limited during the Depression.

During World War II, health insurance was introduced as an employment benefit so that employers could attract and retain skilled workers during the wartime labor shortage. At first, health insurance was inexpensive and there were few organized wellness programs, but interest on the part of government in health education grew following the war. Efforts by employers to build on the opportunities for health gains through worker-oriented programs were fought by organized medicine in the 1950s on the grounds that patients' personal physicians were responsible for their health and well-being, not employers, and as part of the battle against occupational health services on a corporate scale.

In the 1960s a movement took hold called "multiphasic screening" that provided a battery of physical and lab exams to employees on site on an annual basis, the results of which were then passed on to the employee's own physicians. Employers stopped paying for this when they realized that the findings were almost never acted upon, either because they were already known or because the employee's personal physician did not read the report.

The next phase (give or take a few fads) was worksite health promotion. This was a major movement in the late 1970s and through the 1980s. Employers put together educational programs, built fitness centers, changed the menus in cafeterias, and made other modifications that, in the end, were mostly used to advantage by health-conscious, younger workers and rarely by workers who needed them most. A rather large industry developed for the training and certification of fitness trainers, consultants for worksite health promotion, and evaluation methodologists. This resulted in a huge (and overly optimistic because of strong publication bias) literature on the evaluation of these programs. With the downsizing of corporate medical departments around 1980–1985, enthusiasm for these programs waned. Employers, in the end, dumped a lot of these programs in the 1990s but mostly kept the fitness programs because they were a popular employment benefit for headquarters staff. The reason they were dropped is that they were not cost-effective (other than as a competitive enticement for hard-to-find technical employees) and they were not used by the blue-collar workers who needed them most.

In the meantime, however, prevention science had been slowly developing new approaches and methods. More effective means of inventorying and evaluating health risks were developed in the form of “health risk appraisal” instruments. “Behavioral medicine” was a broad movement based loosely on “behaviorism” (a school of thought in psychology that emphasizes “operant conditioning,” reinforcing learned behavior by rewarding or aversive response, and was most closely associated with the psychologist B.F. Skinner). Behavioral medicine began largely outside of mainstream medicine but demonstrated efficacy for a number of clinical modalities such as biofeedback, behavioral therapy for smoking cessation and weight loss, and deeper understanding of psychosomatic disorders. “Health psychology” was a closely related field emphasizing cognitive and emotional responses and their effect on mood and perceptions of health. New approaches to changing health-related behavior on the scale of communities were introduced, particularly “social marketing” (the use of marketing techniques to persuade people to change

behavior). Improved methods for community surveys and the application of more advanced methods of epidemiology to community needs assessment were developed. By the late 1990s, when the field started to slow due to lack of support, prevention science had refined the “technology” available to preventive medicine to measure health needs and to design more effective interventions, both for individuals and communities.

In the mid-1990s, a new movement started called “corporate wellness.” The name is deceiving, however, because the emphasis is not on prevention as much as case management. This movement is basically concerned with demand management, reducing healthcare expenditures, and managing managed care. It was developed earliest in those companies that were self-insured, because they were motivated to reduce costs and had ready access to data (with protection of confidentiality) on individual cases as well as group experience. The movement is also concerned with keeping employees with health problems on the job and productive, for the mutual benefit of both the employer and the worker.

Corporate wellness programs tended to be heavy handed in the early days. In the early 1990s, one major electronics manufacturer and a beer company both proposed to increase the employee contribution to healthcare premiums for employees who did not meet health targets on periodic health evaluations. A well-known manufacturer of candy proposed to fine its employees to recover the increased cost of health insurance for those who did not meet individual health goals and were not on medical treatment plans. These anecdotal reports raised great concern among employees and unnecessary resistance among unions, because of the perception of employer coercion, intrusiveness, discrimination, and compromised privacy.

The corporate wellness movement subsequently moved in another direction, emphasizing optimizing individualized care and managing individual chronic health conditions, such as diabetes, in order to improve quality of life for the worker and to reduce healthcare costs. This trend, poorly documented in the literature, is also led by self-insured employers and seeks to better manage the most

expensive and potentially disabling conditions in the subset of employees at highest risk, most of whom are already under medical treatment and may need help complying with complicated medication regimens and monitoring their condition.

Most recently, as this is being written, wellness and health promotion programs are increasingly motivated by efforts to keep workers productive and on the job for as long as possible. There is a strong social argument to be made that prevention of disability plays a major role in economic security and sustainability. As the population gets older, more people will develop age-related disabilities and chronic illnesses. Unless preventable illness and disability are addressed with strong measures to prevent or delay their onset and progression, the burden of disease and disability will increase and interfere with the ability of more and more people to live vigorous and satisfying lives. The burden of disability and chronic illness will place a rapidly growing burden on the healthcare system and will become a progressively greater social problem if there is no large population growth. Specifically, as the number of people who are retired or cannot work due to disability increases, fewer productive workers contribute to the economic system that supports them, including Social Security, Canada Pension, and healthcare insurance for the aged and disabled (such as Medicare in the United States). This “dependency ratio” has moved and is moving steadily downward in the United States (see Table 19.1). In order to make up this gap, workers will have to become more productive and may (depending on productivity gains) need to work longer and well beyond traditional retirement ages of sixty or sixty-five in order to maintain both national economic stability and their personal income. Disability that interferes with work removes people at a younger age from active employment to dependency and so aggravates the problem.

Some demographers and economists believe that the aging-population aspect of this argument has been overstated, pointing to the British experience, in which labor market changes played a much more important role than aging in skewing the dependency ratio (the arguments are summarized in a book by Mullan, *The Imaginary*

**Table 19.1.** Dependency Ratio for Selected Years

Year	Dependency Ratio*	Historical Note
1950	18:1	Postwar boom period, productivity level low
1965	4:1	Marked increase in productivity per worker due to technology
2005	3:1	Current situation
2080	2:1	Projected to probable retirement age of young people entering workforce today

Source: Economic Policy Institute, EPI Issue Brief #208. Washington DC, 11 May 2008.

*Time Bomb*, 2000), and the observation that American society went through a much greater transition in the dependency ratio in the 1950s and 1960s, which coincided with a period of dramatically increased productivity per worker.

Concern over productivity led the Centers for Disease Control and Prevention, through the National Institute for Occupational Safety and Health (NIOSH), to convene conferences in 2004 and 2007 on wellness and working life that examined these issues. These raised concern among occupational health and safety professionals that NIOSH was being forced to divert energy and resources away from its traditional mission of investigating occupational hazards.

### Periodic Health Evaluation

Periodic health evaluation is the scheduled, routine screening of an individual for health problems when no symptom or obvious health issue has triggered the encounter. Periodic health evaluation for purposes of mandated surveillance is discussed in Chapter 5. Periodic health evaluations for purposes of health maintenance have been extensively reviewed (most recently by Boulware et al. for the American College of Physicians, 2007) and have shown a benefit

with improved delivery of preventive services and reduced concern over health. The periodic health evaluation allows these services to be packaged, programmed with rational periodicity, and incorporated into insurance plans. Piecemeal services do not, and useful preventive services are often omitted when healthcare is limited to occasional physician visits as needed.

Maintaining the health of employees is just as important a commitment of physicians as treating acute injuries and providing clinical care. An employee who is later found to have been unfit for his or her job assignment can be an expensive proposition in terms of liability, bad feelings, and lost productivity. The provision of pre-placement and periodic health evaluations is a responsibility not to be taken lightly. Routine examinations should never become an assembly line, where the doctor just goes through the motions. Physicians who are doing routine examinations should be rotated and the number that they do at any one time should be limited to ensure that they do not get tired, bored, or lose their mental sharpness.

Screening programs are secondary prevention programs organized as campaigns for the detection of treatable early illness. They can feature a single screening modality, such as mammography, or combine modalities into a program. Properly designed, they create opportunities for health education and may include a structured health questionnaire, sometimes called a “health risk assessment,” that identifies serious health risks and may even calculate probabilities of future disease. Referral for follow-up of positive findings and subsequent care is essential in these programs, which are often sponsored by hospitals and healthcare institutions.

“Multiphasic health screening” was a general screening program for large groups of employees that involves a battery of selected, often automated tests given periodically, briefly described above. Multiphasic health screening was often provided for nonexecutive personnel in the form of a mobile screening center, often in a van, that would visit the employer’s facilities in rotation once a year. Because numerous tests were performed in a battery on each participant, and one in twenty was likely to be abnormal, there were usually one or more

abnormal or borderline for every participant. The predictive value of these tests were found to be very low (see Chapter 5), and when the results were sent to the employee's personal physician they were often ignored. As an approach, multiphasic health screening did not emphasize health education opportunities and instead served as a means of conducting screening tests more efficiently. Its effectiveness was often disappointing because the findings, although usually forwarded to the workers' own physician, were seldom actually used in the patient's personal healthcare. The approach is seldom used today.

"Health fairs" are events held at the worksite or in the community under the sponsorship, usually, of a major employer or health-care institution. The health fair is an opportunity for participants to obtain a set of screening tests (such as body mass index, blood pressure determinations, urinalysis by dipstick, spirometry, vision checks, cataract, and inspection of skin lesions for cancer), but tests that require fasting or are difficult to interpret without controlled collection are not recommended, for example serum cholesterol or blood glucose tests. Often free influenza immunizations or other preventive services are offered as well. Although not common at health fairs, cooking demonstrations would be a logical addition and a natural attraction. The health fair is a good opportunity to provide basic health education through free seminars and general advice in a convenient, nonclinical setting, such as booths staffed by nurses and other trained health professionals. The health fair is usually staffed by a nurse with a physician on call or on site but not necessarily involved in every encounter. Some healthcare facilities offer basic health fairs once a year as a free service to employers, in part as a marketing promotion and in part to encourage workers to obtain their own and their family's care at the same facility. For example, UCLA sponsored what it called "HealthFest" on an annual basis for residents of west Los Angeles and their own employees, locating it in an accessible part of their medical center and providing free parking. Child care or entertainment for children is a key element for success of the community health fairs. Prizes and handouts add to the attraction. Health fairs offer timely and nonthreatening opportunities for

health education and probably raise awareness of health issues among community members and workers, at least for a while, but their screening services share the same general drawbacks as multiphasic health screening, unless abnormal findings can be followed up by the sponsoring facility.

The annual periodic health evaluation (more familiarly known as the “annual physical”) is now considered obsolete as a strategy for periodic health evaluation. Most screening tests are not required as often as once a year, especially in younger age groups. The physical examination in itself, even when combined with standard clinical laboratory tests, is now recognized as an inefficient means of detecting health risks or the earliest stages of most treatable disorders. However, the cachet enjoyed by the physical examination after years of promotion by the medical profession and its status as an employment benefit have made it difficult to deny to managers and service employees. Executive health evaluations are discussed below.

The “lifetime health monitoring program” (LHMP) is a phased program that prescribes different screening tests at different frequencies as one ages. The LHMP combines periodic health evaluation with health education and prevention-oriented interventions appropriate to the individual’s stage of life and known risk factors. The LHMP has seldom been offered on a large scale in industry despite its advantages in cost, its efficiency, and its medical rationale. The approach was developed by Lester Breslow and Anne Somers and proposed in 1977 as a cost-effective alternative to the annual physical.

Since 1984, the authoritative guidance for preventive health services in the United States has been the U.S. Preventive Services Task Force (USPSTF), managed by the Agency for Healthcare Research and Quality (AHRQ), Department of Health and Homeland Security. The Task Force disseminates guidelines and updates for evidence-based clinical preventive services intended to assist physician practice in screening and preventing disease. These guidelines cover the spectrum of clinical preventive medicine. Some guidelines are controversial or counterintuitive. For example, USPSTF does not recommend screening for chronic obstructive pulmonary disease by spirometry in

the latest guideline as of this writing (2008) because there is no net benefit in survival or quality of life over and above smoking cessation for smokers and there is a risk of overdiagnosis in “never smokers.” Likewise, USPSTF considers the evidence available to support screening for prostate cancer to be insufficient to make a recommendation for men under seventy-five years of age and sufficient to recommend against it for men over age seventy-five. Because these recommendations are tightly reasoned, are updated periodically, and take into account many factors (among them risk of disease, probable outcome, performance of the screening test, efficacy of the intervention, and risk of adverse effects of treatment), the OEM physician is advised to access the most current guidelines directly from AHRQ and not to depend on printed publications. They are available by category or as a “pocket guide” and as an electronic, downloadable version for PDAs at <http://www.ahrq.gov/clinic/uspstfix.htm>. USPSTF also publishes guides in hard-copy, book form and articles on the rationale and evidence behind its recommendations.

### **Physician Compliance with Preventive Medicine Guidelines**

The prevention-oriented services expected of a physician in a primary care setting are well established. Authoritative guidelines for providing these services are widely disseminated. It is therefore puzzling why physicians do not perform any better than they do in following these guidelines and providing these services of proven efficacy to their patients.

Physicians tend to think that they do better than they do in providing preventive services. There are few recent comprehensive studies of physician compliance with preventive medicine guidelines. Most of what is in the literature dates from the 1980s and 1990s and may not take into account omissions in documentation, but the data overall are not encouraging and probably have not improved much, especially in view of the increasing time demands on the physician.

There are many influences on a physician's practice of prevention. Unfortunately, most current compensation systems do not reward the provision of preventive services to any substantial degree likely to change physician's practice habits. On the contrary, prevention takes time, and physicians are seldom rewarded in quite the same way as they would be if they were to treat a critically ill patient who then recovers. The immediacy of the interaction and feedback is lacking in prevention. This removes the direct satisfaction that a physician may derive from doing well by his or her patient.

Many physicians still carry deep-seeded doubts about the efficacy of preventive services. The treatment of disease can be monitored against an expected outcome (recovery) in the individual patient, but preventive services, if they are successful, result in nothing happening. They can only be monitored on a population basis by demonstrating lower rates of adverse health outcomes, which requires large numbers. Physicians are more motivated to provide preventive services when there is rapid feedback on their performance, test results, and patient outcome.

Training helps to improve compliance with prevention guidelines, at least among family physicians. Also, individuals who trained relatively recently or who are members of professional organizations seem to do better in providing preventive services than those who are relatively isolated from the professional mainstream. OEM physicians, with the advantage of a practice that links individual and population health, are in an excellent position to provide and supervise effective prevention-oriented programs.

### **Executive Health Evaluations**

The executive health evaluation (commonly and colloquially known as the "executive physical") is a periodic comprehensive medical evaluation for executive and other key personnel whose health is thought essential to the employer's future. In actual practice, the annual physical and particularly its executive version continue to be requested by both employees and employers because it is per-

ceived as a perquisite or benefit of employment. One reason is that it provides a convenient opportunity for counseling, health education, and the individualization of health promotion activities. Employers see the executive health evaluation as a means of obtaining an early warning in the event that a company officer whose judgment is critical to the firm may become impaired or unable to carry out his or her duties.

Another reason for their continued popularity is that facilities that provide executive health evaluations are usually located in desirable locations (such as the Mayo Clinic in Scottsdale, AZ, or the Greenbrier Hotel in White Sulphur Springs, WV) and can be combined with short vacations, golf, and other amenities. Most executive health evaluations are very comprehensive and can be very intrusive, raising high expectations for privacy, which make a facility distant from the executive's home office appealing. While confidentiality may be respected throughout the organization, it must be visible to an executive who is used to discretion and deference and who is likely to make snap judgments.

Executive health programs require that the service be scheduled at the convenience of the executive rather than the healthcare provider and that time be managed effectively. In practice this usually means a health interview by questionnaire (usually online), administered before the appointment, and coordinated scheduling in advance of all tests during the visit, such as screening laboratory tests done early in the morning, and expedited access to any necessary diagnostic tests. The visit may be punctuated with breaks for conducting business, checking e-mail, and recreation. The service requires a business arrangement with an upscale hotel nearby that executives will find attractive and that has a full-service business center.

It is highly advisable to create a separate facility for executive health evaluations, preferably on a separate floor of a hospital or in a free-standing location. Combining executive health within general healthcare facility often does not work well. It is not unusual for other patients and for staff to resent the special treatment afforded to executives and for staff to undermine efforts to cultivate the executive

service by allowing delays to occur and refusing to give executives priority over other patients. Also, the creation of a "two-tier" approach to delivering medical services leaves the institution open to criticism for appearing to be discriminating in favor of the affluent and against the needy. Finally, workers who may be directed to occupational health services in the institution may suspect that their cases will be prejudiced or treated less sympathetically because the institution has engaged in an apparently "cozy" relationship with their bosses. These problems can be largely overcome by insisting on as complete a separation as possible between the executive health service and the rest of the organization.

Such programs present an excellent opportunity to encourage sound health practices. The executive participants are likely to be more receptive to concepts of health promotion than they might be at home. Programs on stress reduction, smoking cessation, prevention of alcohol and drug abuse, health education, motivation, and exercise are very popular when combined with screening services. Lifestyle-oriented programs presenting healthier alternatives to the hard-driving executive workaholic stereotype are very attractive in today's more health-conscious society. Spouses of executives can be involved also. Many of these programs can be offered at extra cost or as part of an attractive package within the executive health service.

## **DOMAINS OF PREVENTION**

OEM physicians are members of the preventive medicine and public health communities. The vocabulary of prevention and the nuances of the schools of thought in the field are confusing but important for OEM physicians to know. In part, this is because preventive medicine, like all technical fields, uses jargon as shorthand, and OEM physicians should have no barriers to communication with their colleagues. Jargon is also, as in any field, a means of rapidly sizing up a newcomer. Appropriate and unforced use of the right vocabulary sends a signal to one's colleagues that reassures them with respect to one's expertise. The inappropriate use of words sends a

signal that the user is not prepared and may not be a serious professional and so may close opportunities for cooperation. The definitive guide to the vocabulary of preventive medicine, public health, and epidemiology is the *Dictionary of Epidemiology*, which has been edited for many years by John Last.

The essential first distinction is between “preventive” and “preventative.” Preventive, by traditional usage, not dictionary definition, is the only acceptable word in verbal usage for all applications in this field. “Preventative,” while a synonymous English word in the dictionary, is never used among preventive medicine specialists or prevention professionals. If an OEM physician refers to “preventative medicine,” preventive medicine and public health experts who hear it will immediately discount his or her credibility. The effect is similar to hearing a surgeon refer to the “lar-nix” rather than the “larynx.”

Colleagues outside preventive medicine often use terms such as “preventive medicine,” “public health,” “population medicine,” and “community medicine” interchangeably. However, these terms do have formal definitions, and each of the fields they represent has developed its own culture, tradition, and history among its practitioners.

“Preventive medicine” is defined by Last as “the branch of medicine that is primarily concerned with preventing physical, mental, and emotional disease and injury, in contrast to treating the sick and injured.” He notes further that prevention is often inseparable from treatment and cure, but that the emphasis of the discipline rests on applying many of the same sciences, skills, and attitudes shared by public health to clinical practice. “General preventive medicine” is defined by the American College of Preventive Medicine as “the specialty of preventive medicine which deals with health promotion and disease prevention in communities and in defined populations.”

At one time, “general preventive medicine” was combined with public health in medical specialty boards, and at times they were treated separately. Preventive medicine practice includes population-level interventions as well as services to individuals, which may be educational (usually done by health professionals other than physicians), interventionist (such as immunization), screening for

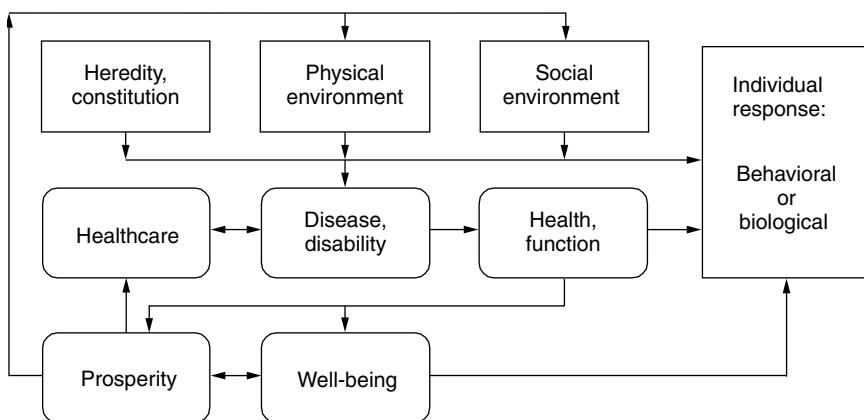
disease, and the evaluation of individual risk, when an individual's entire health profile is considered. Public health, as described in Chapter 8, implies a population- or community-level commitment, such as traditional environmental health services or community-based health promotion programs that touch everybody, rather than individual-level services. Since the two approaches are closely intertwined and usually combined in public agencies, the labels can be considered to refer more to emphasis rather than mutually exclusive missions.

The domain of "clinical preventive medicine," as the term is currently used in the United States, corresponds more closely to the description made by Geoffrey Rose of "preventive medicine" (prevention for individuals and the "high-risk" strategy) compared to public health (prevention for populations and the "population" strategy). A definition of "clinical preventive medicine" was provided in 1982 by a major task force on curriculum development: "those personal health services, provided within the context of clinical medicine, the purpose of which is to maintain health and reduce the risk of disease and untimely death." The definition was seen by many practitioners to be outdated or limiting, and in 1989 a new consensus definition emerged, in which clinical medicine was "an integral part of preventive medicine concerned with the maintenance and promotion of health and the reduction of risk factors that result in injury and disease." Practitioners agree, however, that the domain of the specialty includes behavior modification, risk assessment and modification, health promotion, and health education, all with an individual and clinical emphasis.

"Public health," on the other hand, is defined (again by Last) as "the combination of sciences, skills, and beliefs that are directed to the maintenance and improvement of the health of all the people through collective and social action," and is noted to be oriented primarily to protecting the population concerned, rather than necessarily individuals within the population. An annotated dictionary produced by the U.S. federal government further emphasizes that "public health activities are generally those which are less amenable to being

undertaken or less effective when undertaken on an individual basis, and do not typically include personal health services.” In other words, public health interventions typically involve structural change, as in the formulation of policy and procedures, measures to reduce exposure to hazards, and the assessment of risk.

“Population health” is a relatively new term, emerging in the 1990s and defined in 2002 (Kindle and Stoddart) as “the health outcomes of a group of individuals, including the distribution of such outcomes within the group.” It is not exactly synonymous with public health. It refers to the sum of influences on health from various social and physical factors (the “health field concept”) and the health status of the community as a whole, including distributions, vulnerabilities, and pockets of ill health, as opposed to the health of individuals in the community. A similar usage has become popular in Canada, reflecting the name of a program introduced by the Canadian Institute for Advanced Research that covers studies relating population-based phenomena and social change to health outcomes, irrespective of whether they are targeted public health interventions. (The essential features of this model are presented in Figure 19.1.) “Population medicine” is an older term, also more extensively used in Canada. It was defined by Kerr White, a pioneering thinker in



**Figure 19.1.** The “population health model,” adapted from the Canadian Institute for Advanced Research.

health, and others, as the “application of these concepts and methods embodied in such largely quantitative disciplines as epidemiology, economics, demography, and statistics, and in such behavioral sciences as cultural anthropology, sociology, and social psychology.” Population medicine is similar to its British predecessor “social medicine” in concerns and method, but tends to be more conservative in political tone and liberal (in the philosophical sense of individual-oriented) in orientation than social medicine, which was (and remains) vigorously, enthusiastically, and unabashedly left-wing.

“Community medicine” is much more difficult to define, because its meaning has been expanded in institutional rubrics. The term was used loosely in the 1960s and 1970s as a synonym for preventive medicine, public health, and primary care, with a considerable loss of precision. Its founders defined community medicine as “the academic discipline that deals with the identification and solution of the health problems of communities or human population groups.” The emphasis among purists was always placed on provision of primary care and the elimination of obstacles to obtaining care. Although the ideas are closely related, this is not the same as mobilizing resources specifically to prevent disease and to maintain health in individuals or protecting the health of populations. Community medicine has been described as “a social movement” and “a method for accomplishing social ends.” In practice, the approach used by its founders has been one of community needs assessment, analysis of social and cultural barriers to adequate health services, and provision of primary care in a culturally appropriate structure. The term is out of fashion, but the idea of building community-level health interventions around individual primary care continues in, among other initiatives, the World Health Organization’s drive for primary care for all. The original concept of community medicine is now subsumed under the term “community-oriented primary care.”

Public health may be seen as a collection of disciplines, embracing epidemiology, biostatistics, health services administration, environmental and occupational health, and behavioral sciences, all applied to the protection of populations with respect to health risk. Preventive medicine may be seen as embracing the same disciplines but emphasizing

their clinical application to the individual case. Community medicine may be seen as a systematic approach to all health needs of a defined community. Population health and medicine may be interpreted as perceiving health as an outcome of particular interest given a universe of inputs, outputs, and social constructs. These are not so much four parallel and distinct disciplines as four different approaches transecting the same problem, and inevitably intersecting with one another. When these terms are used interchangeably, important meanings are obscured.

## **HEALTH PROMOTION**

In the 1980s, a new strategy became popular for disease prevention and enhancement of health. This strategy was called health promotion and it assumed the proportions of a social movement toward improved health. The *American Journal of Health Promotion*, which became the leading vehicle for evaluation and scientific studies in health promotion, defined it in 1989 as “the science and art of helping people change their lifestyle to move toward a state of optimal health.” The World Health Organization (WHO) formally defined it in 1986 as “the process of enabling people to increase control over and to improve their health.” Health promotion was not just a repackaged version of clinical prevention, but a strategy emphasizing positive aspects of enhanced health. Instead of adopting the essentially negative approach of “health maintenance,” which underlies the traditional strategies of preventive medicine, the authors of health promotion adopted a more positive approach of “health enhancement,” conveying a new message to participants that not only can disease be prevented and decline arrested, but participants can improve their physical and mental well-being through activities borrowed from clinical medicine, physical fitness, and health education.

There were actually two movements that shared the same name and operated more or less in parallel but as different schools of thought:

- The American school of thought, as represented by Office of the U.S. Surgeon General and the Office of Disease Prevention and Health Promotion, emphasized disease prevention and

fitness and the attainment of explicit goals in population health, codified in a series of goals for the United States that appeared at decadal intervals under the overall title “Healthy People.” (The goals were generally not, in fact met.) The American school of thought has been highly prescriptive, presented largely as a series of rules or measures with which citizens or their physicians should comply.

- The WHO school of thought was embodied in the Ottawa Charter, a consensus document that emphasized social change that would support “healthy public policy,” healthy consumer choices (for example, with respect to food), and individual decisions that lead to healthier life choices (for example, by widespread condom availability). The WHO school of thought has been more influential in the rest of the world.

Health promotion in Canada follows European and WHO models and particularly a commitment to social change, especially modeled on a series of innovations pioneered in the then-health department of the city of Toronto in the 1980s. There, the core concept of health promotion has been to create an environment of choice in which the individual is motivated to pick the healthy choice rather than the least expensive or the most expedient choices in their lives.

Health promotion as it is applied in the United States emphasizes individual initiative and compliance. Health promotion has had particular success in employer-sponsored worksite health promotion programs. Combined with aggressive marketing and backed by employers that recognized the short-term as well as the theoretical long-term advantages, health promotion quickly entered the workplace and became a popular, if unproven, employee benefit in the early 1980s. In recent years an increasing number of employers have introduced health promotion activities for their employees. This movement represents a new point of view complementary to the traditional approach of occupational health services. Health promotion programs based in the workplace or on employment sought to

enhance the personal health of employees for the benefit of both employee and employer. Health promotion was one way in which OEM crossed the line from occupation-related disorders to involvement in non-occupational factors determining health.

Health promotion programs typically blend three approaches to employee health: health education, preventive medicine and health screening, and fitness programs.

The health education component is concerned with teaching employees the essentials of a healthy lifestyle, such as good health habits, sound nutrition, and the consequences of smoking, alcohol, and drug abuse. Beyond the informative aspect of health education, however, is attention to the psychological principles that motivate people to comply with sound health practices or to take unnecessary risks that jeopardize their health. Simple information transfer is not enough, which is why lectures on health seldom change behavior. Likewise approaches that may be appropriate to patient education are not always helpful in health education for well employees. A health education component within a workplace health promotion program must be designed with attention to the characteristics of the population of workers to be reached (age, sex, class, education, health status, language, etc.), the most important health problems in the community as perceived by the workers, the most important health problems actually present in the community, and the goals to be attained in changing health-related behavior. Programs that fail to address these basic concerns are usually ineffective. Meeting these needs usually requires professional consultation and is beyond the training and experience of most physicians.

The preventive medicine component is typically limited to screening for common disorders and risk factors and intervention activities that complement but do not substitute for personal health-care, such as smoking cessation. The emphasis in health promotion programs is on primary prevention to reduce disease incidence in the working population and secondary prevention, the early detection of disease and referral for care. Reducing risk factors for later health problems (such as reducing cholesterol levels of cardiovascular fitness

training) is more easily accomplished in an integrated health promotion program. In such programs the group spirit, constant encouragement and feedback, and support network make compliance easier to achieve than on an individual education, preventive medicine, and physical conditioning.

Health education is an integral part of these programs. For community and employee-sponsored programs, health education opportunities may be passive dissemination of information, through newsletters or emails, or active, with discussion groups and employee feedback on topics workers would like covered. “Lunchtime listen and learn” seminars, with an invited speaker available over the lunchtime break, are particularly popular and can be shared among employers in the same building.

Physical conditioning is an important part of most health promotion programs for several reasons. Participation in fitness activities and sports programs provides a regularly scheduled opportunity for conditioning with encouragement and the sense of belonging to a group, both of which make compliance more likely. Fitness programs within worksite health promotion are discussed below.

The concept of health promotion is a powerful one because it places a positive emphasis on enhanced well-being as a social expectation. As a strategy, health promotion gives the momentum to the proponents of sound health-related behaviors and to the advocates of constructive social change. It gives the individual participant a sense of belonging to a larger community of health-conscious peers. One would expect this strategy to be more effective than trying to persuade people to cease practicing unhealthy habits that they may enjoy or that give them a sense of comfort. As nominally healthful behaviors become a social norm, increasing pressure is placed on those who do not comply for reasons of personal choice or the appropriateness of the recommended behavior to their needs and abilities. This is evident in contemporary social attitudes toward smoking and obesity, for example, which suggest that the health messages of the previous decade were internalized into both individual thinking and the culture.

“Corporate culture” programs are highly sophisticated campaigns that link the employers’ image and identification with health and vitality. The emphasis is on energy, optimism, productivity, and teamwork. These campaigns weave health promotion elements throughout, including fitness programs, fundraising for health-related appeals, heart healthy cafeteria menus, and positive thinking.

Whatever the model of health promotion program, there are certain implicit principles that should be followed. Effective programs are easily accessible to employees, both geographically and culturally. They are adaptable to individual needs and emphasize personal responsibility and empowerment. They may contribute to a positive corporate culture and image, but they are not extensions of human resources departments or ways of monitoring employee behavior outside of work hours. It is never acceptable to break confidentiality with personal health information. Employees who choose not to participate in designated activities such as fitness programs should not be coerced or made to feel that they are letting down “the team.” For this reason it is not appropriate to combine health promotion programs with athletic competition.

### **Worksite Health Promotion Programs**

Employer-sponsored health promotion programs may be community-based, such as the provision of discounted gym memberships, or worksite-based. Worksite health promotion began in the 1970s and became a major movement in the 1980s. In the beginning of this movement, worksite health promotion programs were usually started and managed by trainers and trained program directors, some of whom were certified by one of several credentialing agencies that sought to professionalize the position. The role of the primary care physician in worksite health promotion programs was usually that of an adviser. In many cases, OEM physicians who served as corporate medical directors were reluctant to get their departments involved in worksite health promotion because they could not see the sustainable health benefit and feared that their energies would be detracted from

more pressing issues of occupational hazard control. During the downsizing and delayering that followed, however, health promotion programs were often assigned to them anyway and in time became a part of the corporate medical department's structure. Many physicians enjoy the opportunity to encourage good health habits in an upbeat situation that these programs provide.

Worksite health promotion has been supported by a very large, enthusiastic commercial sector that advocates and markets powerfully for health promotion services that they would supply or manage. As is the case for health services generally, there is great interest in the private sector in franchising worksite health promotion services. This is particularly true because reimbursement for preventive medical services has been depressed but has the potential to rebound in mainstream medicine, as demonstrated by the public's huge out-of-pocket expenditures for prevention-oriented alternative care and nutrition. But in many companies, health promotion programs were seen as expensive nonproductive costs and were terminated during the 1990s.

Worksite health promotion programs may be based on company grounds or they may operate by encouraging or subsidizing employees to use a community facility. Some are directly run by the employer, but many of the most successful programs are sponsored through employee recreation associations. Management of such programs has become highly specialized and professional, but small-scale programs can begin on an inexpensive basis. Worksite or employer-sponsored health promotion programs offer an opportunity to link health promotion, cost containment, and self-management of health-care. Self-management will require a considerable educational effort to change people's perceptions of what they should do for themselves, when they should seek medical care, and whom they should go to when they need care. The worker's personal physician will be in a key position to support these changes and to reinforce that education.

Worksite health promotion programs target those health-related behaviors that are most appropriate to employed persons, who are

generally between twenty and sixty-five years old and in good health. They tend to attract more youthful, physically fit, and middle-class workers who are already aware of health risk factors. Blue-collar workers, linguistically or ethnically isolated workers, workers in rural areas, or workers whose worksite is remote from headquarters tend not to participate, although there is evidence that they stand to benefit as much or more than white-collar, urban workers. Executives tend to prefer separate programs, but for ensuring motivation and credibility among employees it is also important that the executives be seen to participate in worksite programs themselves.

Worksite health promotion programs may play an extremely important and constructive part in an individual worker's healthcare if the worker is motivated to participate. It is a solid adjunct to personal prevention at home and a healthy lifestyle. The worksite program offers an opportunity to target individual risk factors. For example, a worksite health promotion program may be the most appropriate means for prescribed exercise programs, for nutritional counseling, or for support in lifestyle changes such as smoking cessation or weight reduction. In order for the physician to take advantage of this opportunity, the worker-patient and his or her personal physician should discuss the program and what opportunities are available.

Most worksite health promotion programs are not supervised directly by a physician on site, although many have physicians as consultants or advisors. When there is a special need or opportunity, the worker's physician should not hesitate to contact the physician associated with the program in order to maximize the benefit and to minimize the risk of that worker participating in the program.

Evaluation of the impact of worksite health promotion programs suggests that well-managed worksite health promotion programs succeed in reducing illness and healthcare utilization, improving employee morale, and encouraging lifelong good health habits. Self-selection and attrition remain major limitations to the evaluation of worksite health promotion programs. However, effectiveness is not the only reason for their popularity, and evaluation of effectiveness is not the only criteria used by management in deciding to support

such programs. The programs are very popular among employees, so most organizations wish to retain them. Thus, like health insurance after World War II, it has been difficult for many employers to phase out worksite health promotion programs once they started, regardless of whether they are deemed to be successful. Some employers saw no reason to evaluate them, since their continued presence served a business need in retaining employees.

Companies that are more committed to health promotion tend to be larger and involved in high technology, in which case they are usually in competition for desirable employees, who tend to be younger, well-educated, and health conscious. The employees drawn to worksite health promotion programs therefore tend to be concerned with their health and already motivated. More women than men participate, and women tend to participate in more activities, particularly those that involve interpersonal skills, stress reduction, and weight control. Sociological theory and hard-nosed management experience both confirm that the motivating force behind the introduction of worksite health promotion programs is not cost effectiveness or the control of healthcare costs. In the view of many, it is the social importance of health promotion as a means for organizations to attract and hold desirable employees, to express their concern for employee's well-being, and to offset problematical cutbacks in benefits while maintaining the morale of the workforce.

Health promotion programs continue to have less visible influence among rural, working-class, and socially isolated subgroups than with urban, better educated, and more affluent North Americans. This class distinction may ultimately disappear due to social mobility and marketing.

### **Elements of Worksite Health Promotion Programs**

Certain services and interventions are common and important as a part of worksite programs, as listed in Table 19.2.

Fitness programs are by far the most popular program elements and are seen by a substantial fraction of employees—usually younger and

**Table 19.2.** Typical Components of Worksite Health Promotion Programs

<i>Health Education</i>	
Cancer prevention	Common minor illnesses
Heart disease	Child health
Mental health	Care of the elderly
Nutrition	Diabetes
Substance abuse	Allergies
Smoking	Automotive safety
Injury prevention	Families relations
Breast self-examination	Travel medicine
Back school (prophylactic)	Sports
Care of dependents	Aging
Hair loss	Cosmetic surgery
<i>Preventive Medicine</i>	
<i>Screening activities</i>	<i>Intervention activities</i>
Hypertension screening	Smoking cessation
Diabetes screening	Dietary interventions
Glaucoma screening	Back school (rehabilitation)
Cardiovascular risk factors	Weight control
Pulmonary function testing	Stress reduction
Weight monitoring	Prescriptive exercise regimes
Breast examination	Unsupervised exercise
Stool occult blood	Hypertension control, monitoring
<i>Personal Wellness and Self-Care</i>	
Anxiety disorders	
Arthritis	
Asthma	
Chronic pain	
Diabetes	
Gastroesophageal reflux disorder	
Headache	
Hypertension	
Osteoporosis	
Pregnancy	
Urinary tract infections	

already fit—as a major employment benefit. Access to fitness facilities onsite or nearby in the workers' own home community is highly attractive to workers who are already persuaded of the benefits of fitness. It is also common for employees to organize their own informal fitness programs, such as lunchtime aerobics, with management cooperation. In the past two decades most large companies have enthusiastically joined this trend, often installing well-equipped fitness facilities at their headquarters and major locations and hiring trainers and coordinators to staff them. These facilities are very expensive and are cost-effective only when large numbers of employees are concentrated in one place. An alternative is for the employer to pay for memberships in local fitness clubs or gyms, although the identification with the employer tends to be less strong in this case. A key to the success of fitness programs beyond lunchtime aerobics is to make available a variety of levels of involvement so that workers can individualize their regimens according to need, preference, and personal schedules. Another key to the success of fitness programs is the availability of a sufficient number of showers. Fitness programs can be started on a shoestring as lunchtime aerobics and stepped up from there. Men, particularly, often see the facilities as an opportunity to regain or maintain athletic levels of fitness, but an emphasis on this level of commitment , and the monopoly of fitness facilities by hard-bodied regulars, may discourage participation by out-of-shape workers, who need the facilities the most. It is important that fitness programs encourage participation at all levels and discourage attitudes that promote a "jock" or "gym" mentality of personal competition, cliquishness, and singles cruising. Properly supervised, fitness-activity related injuries are much less common than unsupervised sports-related injuries and on balance appear to reduce the risk of workplace injuries among participating workers, although this has been difficult to document. Since workplace fitness programs are not oriented to athletes, persons of normal strength and coordination can set reasonable goals and attain them; the emphasis should not be on high performance. Fitness programs are fun and build morale among employees. From the point of view of management, fitness programs

are excellent opportunities to build morale and team spirit among employees.

Nutrition and weight control are very popular program elements that tap into the concern of North Americans about their weight and eating habits. Because weight reduction is best achieved through reinforcement of diet and exercise behavior, it makes sense that group programs would be effective, although this remains unproven for the long term. Further, because objective nutritional education is hard to come by in daily life in a society devoted to product marketing, the nutritional counseling component of such programs is very popular. Weight control programs dovetail well with other health promotion activities, particularly fitness and cardiovascular risk factor modification, and lend themselves well to campaigns and institutional changes (particularly light cafeteria menus). They are good ways to start or to introduce a more comprehensive health promotion program in the workplace. One approach has been an incentive in which the employer makes a contribution to an employee's individual "cafeteria fund" (i.e., the company helps to buy lunch, but it has to be healthy), which demonstrates that simple measures to achieve gains are possible within the health promotion paradigm.

Smoking cessation programs are common, popular, and often effective, especially when combined with a change in corporate smoking policies that reduce opportunities to smoke. Help to stop smoking is very nearly an ethical necessity when smoking is abruptly prohibited in a facility or workplace. Because approximately one-third of smokers are thought to be sufficiently addicted to cigarettes to be unable or unwilling to quit on their own, corporate smoking cessation programs can provide a real community service.

Cardiovascular risk factor modification programs include all of the elements previously mentioned as well as regular cholesterol and blood pressure determinations. Managing individuals with clinical hypertension or major elevations in blood cholesterol is primarily a medical responsibility and is beyond the scope of worksite health promotion programs. However, such programs may be an opportunity for on-site medical facilities to monitor health at work for

employees at risk. Cardiovascular risk factor modification in worksite health promotion programs plays a supporting role in modifying the health risk of the great majority of persons with normal blood lipid phenotypes who can control their level of risk within the “non-high-risk” group by simple dietary adjustments and exercise.

Stress management is a popular theme in health promotion programs and is sometimes introduced during critical transitions in the company, especially during times of layoffs. The topic is discussed in Chapter 13.

## EMPLOYEE ASSISTANCE

“Employee assistance” is a term, adopted in 1974 by the National Institute on Alcoholism and Alcohol Abuse, that refers to an organized system to provide assistance to employees with personal problems, mostly related to substance abuse. Employee assistance is by definition a form of secondary and sometimes tertiary prevention, because it is designed to identify and stop the progression of an existing health problem and to assist the worker in obtaining skills and building the motivation to overcome it. The guiding philosophy of employee assistance is that the worker is more likely to recover and manage an addictive behavior or to break other health-adverse habits with the support of a community and when something the person values is at stake, like a job. Employee assistance is a way of helping troubled employees for their own sake, giving workers with problems a second chance, and saving value for the employer by retaining skilled and loyal employees.

Employee assistance programs (EAPs) identify workers with personal problems, refer them for treatment, support and motivate them to complete treatment, and assist in their rehabilitation. The majority of “broad brush” EAPs are really substance abuse and financial counseling programs with a relatively small primary preventive aspect, usually in the form of low-key health education. Participants in EAPs either self-refer or are referred to the EAP by management because of performance problems or positive drug screening tests.

(Drug screening is discussed in Chapter 18 because it pertains primarily to fitness for duty.) Most EAPs are focused on alcohol and drug abuse and mental illness, but many concern themselves with family and adjustment problems, financial mismanagement (particularly credit card overruns), grief management, adjustment crises (such as helping victims of crime or loss due to a disaster such as a fire), job burnout, stress in and outside the workplace, domestic violence, and mental illness. EAPs can be pivotal in the management of mental disorders, which is the second major category of disability in the United States and among the most costly diagnostic category for employer-sponsored health insurance, primarily because of long-term disability and duration of treatment and absence incidents. An important role of EAPs is helping the employee to reestablish balance in work and life.

EAPs usually do not provide on-site counseling and therapy beyond an initial session or crisis intervention. They identify workers with personal problems, refer them for treatment, monitor their progress, support and motivate them to complete treatment, and assist in their rehabilitation. The EAP also helps to determine fitness to return to work (see Chapter 18) and assists in the reintegration of workers.

EAPs are a very mature intervention strategy, with precedents going back decades, established protocols, and a wealth of empirical experience. The literature is summarized in one authoritative source, making this one of the few fields of public health that, uniquely, can be encapsulated in a single handbook, a work by Oher that has been through many editions and is considered the standard reference in the field.

An EAP operates primarily by self-referral of patients, who are then referred to local healthcare or counseling facilities. Some workers are sent to EAP programs by their supervisors as a condition of retaining employment when their job performance has suffered or they have appeared to be impaired. The employer is informed of the progress of the employee's rehabilitation and guarantees return to the same or similar work when recovery is sufficient. Confidential

information, such as diagnosis, treatment, and the content of interviews, is not shared with management. EAPs usually do not provide direct treatment except for initial counseling. Instead, EAPs usually rely on existing community services.

If a worker presents signs of personal problems, anxiety, or substance abuse, that worker may be helped by self-referral to an employer's EAP, if one is available. In the absence of an EAP, the physician can perform the initial evaluation, triage, and, in some cases, begin treatment on an individual basis. Many employers will be cooperative and may assume costs for key or long-term employees if reimbursement under the health plan is not complete. It is not unusual for the cost of treatment to be shared between the employer and the health plan or a private insurer, depending on local arrangements. Maintaining patient confidentiality is essential at every step.

The ideal is of course voluntary referral by individuals who recognize that they may have a problem affecting their life and work. Obviously, some insight and honesty are required to recognize the symptoms, and people in denial may be incapable of recognizing the cardinal symptoms of alcohol abuse.

The emphasis should always be on voluntary referrals, but acceptance of mandatory referral may be necessary if the alternative is for the individual to be fired after a certain number of warnings. One cannot count on such people having the insight to refer themselves, however, and for that reason mandatory referrals are almost inevitable if a program is to be truly effective as an alternative to firing the impaired worker. Mandatory referral should only be used as a last resort, however, and every effort should be made to encourage voluntary referral, at least in part because the programs work better for those who have taken the initiative to refer themselves. This means promoting the program heavily and insuring that confidentiality is kept strictly. Privacy measures should be highly visible; for example, the EAP office should be off site and away from view of other employees. (Two hospitals on opposite sides of the island of Oahu, Hawaii, for example, each with EAP programs, had an agreement to take one another's employees.)

As a general rule, if an EAP program is operating primarily with voluntary referrals, it is doing well. If most of the participants are there because of mandatory referrals due to poor work performance, this suggests a problem with how the EAP is perceived. The program may not be trusted by employees, it may be overused by supervisors, it may not be sufficiently promoted, or the workers for whom it is intended may not view it as a viable resource for their problems.

The supervisor's role in an EAP should be kept limited but is essential to the process. The supervisor may play a very constructive role in identifying the problem since supervisors are often the first to see a decline in work performance or other indications that a worker is in trouble. The supervisor may also be able to arrange the worker's schedule and duties to accommodate participation in the program and to ensure easy return on completion. The supervisor, in short, cannot be kept out of the process but should be involved only to a limited extent in keeping with the confidentiality of medical information. The supervisor has no business receiving any sort of medical or personal information on the worker, although it is impossible to stop rumors and information through the shop grapevine reaching him or her. The supervisor is entitled to receive regular reports on the worker's fitness to return to work and to assume the duties required by the job. The supervisor does not need to receive any specific diagnosis or sensitive information, only whether or not the individual can do the job and if the individual cannot what the individual is capable of doing if there is evidence of impairment. The supervisor should also be told if there is any possibility that the individual could be a risk to others. In a humane and supportive working environment, the supervisor can be a big help in facilitating the return to normal working life of the recovering worker. It is important, however, for the worker, the supervisor, the company, and the union to be entirely clear on the ground rules ahead of time to prevent problems. There is obviously a fine line between a supervisor's human concern on the one side and meddling and paternalism on the other.

EAPs appear to be cost-effective for the employer, particularly in large organizations, depending on its structure and level of activity. It

is not unusual for the cost of treatment to be shared between the employer and the health plan or a private insurer, depending on local arrangements.

There is a sharp distinction to be made between employee assistance programs and health promotion programs. The two types of programs should be kept distinct and under separate organizations. Employee assistance programs are very sensitive and should be tightly controlled, while health promotion programs should be more relaxed and activities are usually best done in groups.

Referral to an outside counselor or health professional for evaluation and treatment, rather than treatment by an internal unit within the organization, is generally preferable. Unless an employer is very large, it is difficult to preserve the confidentiality required when workers receive their care at the worksite. The employee assistance counselor should be isolated from the pressures of management (as, ideally, should the occupational physician), but in practice it is very difficult to stand alone in the organization against requests for information and certification to return an employee to work. Many effective programs rely on external agencies to provide treatment and medical or psychological evaluations and use their in-house counselor to monitor progress through regular reporting and discussions with the treating health professionals or for crisis interventions when necessary. This approach is also more cost effective, in general, than establishing an elaborate treatment program onsite. Having the evaluation and treatment at arm's length from the employer keeps a degree of objectivity, removes the suspicion that management might interfere, and reduces mistrust over the quality of care being given.

## **DISEASE/CASE MANAGEMENT**

Employers have become increasingly interested in managing high-cost cases through assistance in scheduling, monitoring compliance, referrals to specialized care, convenient workplace health monitoring (through their occupational health services), education and behavioral medicine, pharmacy-care programs and tertiary prevention

(interventions to prevent disabilities and disease progression). This trend, which is growing in strength in the business community, follows the observation that individualized risk factor intervention for high-risk employees results in more favorable outcomes than broad employee health promotion programs alone.

An example of the disease management approach is a program introduced by Lucent Technologies. Employees were screened for cardiovascular risk factors, and qualified high-risk employees were then supported through exercise/fitness programs, educational programs, dietary change, and individualized on-site counseling in the workplace. The program achieved a high level of employee satisfaction, identified 2.4 percent of the employees as having diabetes, and resulted in 17 percent of the employees beginning cardiovascular medication. On the other hand, in one employer-sponsored cardiovascular risk reduction program, employees who were followed up with structured programs after screening did not do as well as those who chose informal means of health risk management. Thus, there is much that remains to be defined and clarified in the disease/case management approach.

The principal conditions that major employers considered to be priorities (over 40 percent deemed highly important) are those that incur the greatest costs in productivity and health-related expenses: back pain, musculoskeletal disorders, depression and other mental disorders, repetitive strain injury, cardiovascular risk factors (hypertension, obesity and diabetes), substance abuse, smoking-related problems, and influenza. Arthritis and headache/migraine are also known to be associated with a disproportionate loss of productivity. However, employers rated their performance in managing most of these priority conditions as deficient. The conclusion to be drawn from these and other data is that there is a large performance gap in management of disorders connected with the greatest loss of productivity and that current models of intervention and behavior change are not meeting the need.

The movement for intensive case management is not well documented in the literature, unlike the abundant literature on the health

promotion movement. Employers are engaging in this approach reluctantly, aware that it may be considered intrusive by employees and unions, and many employers believe that they are being forced, in effect, to assume responsibility for direct delivery of care, which is outside their core business and comfort zone.

Not as developed but equally important with respect to social benefit is the measure of the advantages of workplace-centered wellness programs to workers and their families: improved general health and vitality, reduced risk of catastrophic illness, enhanced employment security, enhanced productivity in non-employment-related activities, and protected social capacity, by which is meant the ability to play social roles as, for example, a parent, friend, community leader or civic participant.

## **WELLNESS AND PRODUCTIVITY MANAGEMENT**

Employers today are confronted by severely rising costs and declining economic prospects, especially during the current recession, which have led employer-payers to consider new healthcare reform strategies. The strategy of wellness and productivity management emphasizes prevention, managed care, and marketplace mechanisms and has increasingly considered previously unacceptable measures that involve intensive case management, individual intervention, and personal tracking that would previously have been considered overly intrusive on the part of an employer. However, faced with the burden of paying for healthcare coverage, employers are increasingly assertive in their role as healthcare managers and see no alternative to managing their employees' health. What is new about this approach (which is not well documented in the medical literature) is that efforts are being more tightly targeted and the interventions are specifically designed to reduce high-cost outcomes such as chronic disability.

This framework emphasizes science-based health and human performance interventions intended to target opportunities for controlling costs and enhancing productivity. This would be achieved

through reduced sickness incidents, improved case outcomes, decreased “presenteeism” (the presence of a worker on the job but not working at full capacity, due to low-grade illness or incapacity), and enhanced physical capacity to do the job. Most cardiovascular and mental disorders are the leading targets, and the intervention may take the form of health promotion programs or intensive case management for chronic disorders associated with high cost and the risk of disability, such as diabetes.

The prevailing philosophy of this strategy has been that health-care costs could be brought under control, or at least reduced into more manageable proportions, by reducing both the actual need and the demand for market-driven health services. This strategy encourages interventions targeted to prevent loss of productive years of life and to prevent disability, rather than the more traditional goals of extending life and preventing disease. For example, one major initiative undertaken by Thomas Jefferson University was to quantify lost productivity due to migraine headaches, on the theory that intensive management could result in considerable cost savings that would recover the cost of introducing and maintaining the program. This policy approach may require substantial cultural change to discourage risk-taking behavior, to devolve responsibility for personal health management (and triage) on the individual from a practitioner or system, and to manage individual cases through intensive case management, facilitated scheduling, or services and provider discounts (such as drug plans). The goals would be achieved by, respectively, health promotion (and self-care), managed care, and intensive case management.

As the attention of industry has shifted from cost control and loss reduction to productivity, the tools for measuring performance and assessing efficacy have also changed and have become much more sophisticated. Disease outcomes and the cost of disability are relatively crude measures. More sensitive financial and behavioral indicators have been developed to assess productivity on a micro level, for evaluation purposes, and to identify opportunities for intervention. These include the Health and Labor Questionnaire, the Work

Productivity and Activity Impairment Questionnaire, the Osterhaus productivity technique (based on frequency of presence and absence and developed for migraine studies), the MacArthur Health and Performance Questionnaire, the Work Limitation Questionnaire, and the Stanford Presenteeism Scale, among many others. The measures used to do this include time on task in the workplace, work quality, work quantity (productivity), interpersonal functioning in the workplace, and work culture. A comprehensive toolkit is now available from the American College of Occupational and Environmental Medicine.

An important aspect of productivity research is the financial gain to the employer of promoting health among employees. Current methodology stresses identifying the “break-even” point at which an investment in wellness covers the cost of operating a program. However, senior managers in industry are oriented more toward comparing alternative rates of return than either to loss reduction or covering costs alone. This strategy may backfire if unscrupulous managers use tactics to exclude high-risk individuals from the workforce or to make jobs more difficult for the disabled (fortunately, such tactics have been made more difficult with the requirements of the Americans with Disabilities Act) rather than modifying the work or introducing wellness programs. The OEM physician often faces a challenge in influencing senior management to do the right thing, but a good business case combined with ethics and a corporate policy spelling out commitment can be very persuasive.

## **RESOURCES**

- Boulware LE, Marinopoulos S, Phillips KA, Hwang CW, Maynor K, Merenstein D, Wilson RF, Barnes GJ, Bass EB, Powe NR, Daumit GL. The value of the periodic health evaluation (systematic review). *Ann Intern Med.* 2007;146(4): 289–300.
- Swotinsky RB, Smith DR. *The Medical Review Officer's Manual: MROCC's Guide to Drug Testing.* Beverley Farms, MA: OEM Press; 2006.
- U.S. Preventive Services Task Force. *The Guide to Clinical Preventive Services.* Washington DC, Agency for Healthcare Research and Management, updated annually. Available at [www.ahrq.gov/clinic/uspstf/usptopics.htm](http://www.ahrq.gov/clinic/uspstf/usptopics.htm).

## NOTEWORTHY READINGS

- American College of Occupational and Environmental Medicine, ACOEM Health-Related Productivity Roundtable. *The Health of the Workforce and its Impact on Business*. Arlington Heights, IL: ACOEM; 2003.
- American College of Occupational and Environmental Medicine, and The Benfield Group. *Strategic Management Needs in Health and Productivity*. Arlington Heights, IL: ACOEM; 2003.
- American College of Occupational and Environmental Medicine, and the Office of Health Policy and Clinical Outcomes, Thomas Jefferson University. *Health-Related Work Loss and Productivity*. Arlington Heights, IL: ACOEM; 2002.
- American College of Physicians, Medical Practice Committee. Periodic health examination: a guide for designing individualized preventive health care in the asymptomatic employee. *Ann Intern Med.* 1981;95:729-732.
- Battista RN, Lawrence RS. Implementing Preventive Services. *Am J Prev Med.* 1988 (suppl);4(4).
- Bertera RL. Planning and implementing health promotion in the workplace: a case study of the Du Pont company experience. *Health Educ Q.* 1990;17:307-317.
- Bertera RL. The effects of workplace health promotion on absenteeism and employment costs in a large industrial population. *Am J Pub Health.* 1990;80:1101-1105.
- Borland R, et al. Changes in acceptance of workplace smoking bans following their implementation: a prospective study. *Prev Med.* 1990;19:314-322.
- Brady W, Bass J, Moser R, Anstadt G, Loepke R, Leopold R. Total corporate health care costs. *J Occup Environ Med.* 1997;39:224-231.
- Breslow L, Somers AR. The lifetime health monitoring program—a practical approach to preventive medicine. *N Engl J Med.* 1977;22:588-592.
- Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J.* 1979;12:1194-1254.
- Canadian Task Force on the Periodic Health Examination. The periodic health examination: 1. 1984 update. *Can Med Assoc J.* 1984;130:1278-1292.
- Charap MH. The periodic health examination: genesis of a myth. *Ann Intern Med.* 1981;95:733-735.
- Conrad P. Worksite health promotion (special issue). *Social Sci Med.* 1988;26(5):485-575.
- Goldbloom R, Battista RN, Haggerty J. The periodic health examination: 1. Introduction, 2. 1989 update. *Can Med Assoc J.* 1989;141:205-207, 209-216.
- Green H. *Fit for America: Health, Fitness, Sport and American Society*. New York: Pantheon Books; 1986.
- Gross PA, Cataruzolo P, Mitofsky W, Furnari M, Crupi T, Skurnick JH, DeMauro P, Statmore G, Moagan M, Berdy J, Sokol C. Implementing

- preventive health measures: a pilot study. *Clin Perform Qual Health Care.* 1999;7(2):52–55.
- Guico-Pabio CJ, Cioffi L, Shoner LG. The Lucent-takes-heart cardiovascular health management program: successful workplace screening. *AAOHN J.* 2002;50:365–372.
- Guidotti TL. Adaptation of the lifetime health monitoring concept to defined employee groups not at exceptional risk. *J Occup Med.* 1983;25:731–736.
- Guidotti TL. Comparative models for workplace-centered wellness. *J Health Productivity.* 2008;3(1):15–21.
- Han PKJ. Historical changes in the objectives of the periodic health examination. *Ann Intern Med.* 1997;127:910–917.
- Health Promotion Research Methods: Expanding the Repertoire.* Conference overview and selected papers. *Can J Public Health.* 1992 (suppl 1); 83.
- Health Promotion: A Discussion Document on the Concept and Principles.* Copenhagen: World Health Organization Regional Office for Europe, ICP/HSR 602, 1984.
- Hopkins WG, Walker NP. The meaning of “physical fitness”. *Prev Med.* 1988;17:764–773.
- Keating DP, Hertzman C, eds. *Developmental Health and the Wealth of Nations: Social, Biological and Educational Dynamics.* London: Guildford; 1999.
- Labonté R. Population health and health promotion: what do they have to say to each other? *Can J Pub Health.* 1995;86:165–168.
- Lalonde M, Health and Welfare Canada. *A New Perspective on the Health of Canadians.* Ottawa: Health and Welfare Canada; 1975. [The Lalonde Report]
- Mullan P. *The Imaginary Time Bomb.* London: I. B. Tauris; 2000.
- Ockene JK, Edgerton EA, Teutsch SM, et al. Integrating evidence-based clinical and community strategies to improve health. *Am J Prev Med.* 2007;32:244–252.
- Oher JM. *The Employee Assistance Handbook.* New York: Wiley; 1999.
- Ozminkowski RJ, Goetzel RZ, Santoro J, Saenz B, Eley C, Gorsley B. Estimating risk reduction required to break even in a health promotion program. *Am J Health Promo.* 2004;18:316–325.
- Pelletier KR. A review and analysis of the clinical- and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: 1998–2000 update. *Am J Health Promo.* 2001;16:107–116.
- Pescatello LS, Murphy D, Vollono J, Lynch E, Bernene J, Costanzo D. The cardiovascular health impact of an incentive worksite health promotion program. *Am J Health Promo.* 2001;16:16–20.
- Pollmar TC, Brandt ENJ, Baird MA. Health and behavior: the interplay of biological, behavioral, and social influences. Summary of an institute of medicine report. *Am J Health Promo.* 2002;16:206–219.

- Russell LB. The role of prevention in health reform. *N Engl J Med.* 1993;329: 352–354.
- Sagan LA. *The Health of Nations: True Causes of Sickness and Well-Being*. New York: Basic Books; 1987.
- Schneider WJ, Stewart SC, Haughey M. Health promotion in a scheduled cyclical format. *J Occup Med.* 1989;31:443–446.
- Sciacca J, Seehafer R, Reed R, Mulvaney D. The impact of participation in health promotion on medical costs: a reconsideration of the Blue Cross Blue Shield of Indiana study. *Am J Health Promo.* 1993;7:374–383.
- Scofield ME, ed. Worksite health promotion. *Occup Med.* 1990;5(4):653–876.
- Self-Care Institute. *Demand Management Handbook*. Washington, DC: Partnership for Prevention; 1996.
- Stein AD, Shakour SK, Zuidema RA. Financial incentives, participation in employer-sponsored health promotion and changes in employee health and productivity: HealthPlus health quotient program. *J Occup Environ Med.* 2000;42:1148–1155.
- Terris M. The distinction between public health and community/social/ preventive medicine. *J Pub Health Pol.* 1985;435–439.
- World Health Organization. *The Ottawa Charter for Health Promotion*. Ottawa: World Health Organization, Health and Welfare Canada, and Canadian Public Health Association; 1986.

*This page intentionally left blank*

# 20 OCCUPATIONAL HEALTH SERVICES

An occupational health service (OHS) is a facility or unit that provides healthcare and prevention services to workers. In this chapter, the term is used to refer to staff and facilities providing medical care, with or without support services. An OHS in this sense is analogous to a “service” in a hospital. (For some reason, OHSs are rarely called “occupational medicine services.”) OHSs normally consist of an occupational medicine clinic and supporting services for hazard control, wellness, and evaluation. They may be inside the organization or outside, providing services on contract or case by case.

OHSs have a medical orientation and provide healthcare services. The units called “occupational health and safety departments,” on the other hand, manage workplace health protection but do not provide healthcare. Occupational health and safety departments are within the organization but are not necessarily comprehensive; they may contract out for services such as occupational (industrial) hygiene.

Every OHS has its own mission and reporting relationships. In general, the purposes of an OHS are to:

- Manage injuries, illness, and disability resulting from work
- Protect employees from occupational risks where they work

- Prevent disability by providing or managing the highest level of care followed by support during recovery and rehabilitation as needed
- Achieve the highest practicable level of health and productivity among all employees of the enterprise for which it is responsible
- Ensure that work capacity, production requirements, and job requirements match
- Support the business, that is, assist in the management of workers' compensation, insurance, administration, monitoring, and productivity

This chapter will discuss the role of various types of OHSs, their organization, their operation, and their limitations.

## **THE OCCUPATIONAL HEALTHCARE SYSTEM**

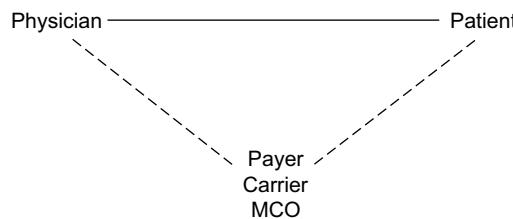
The occupational healthcare system exists in parallel with the larger general healthcare system in North America. The general healthcare system is undergoing rapid and fundamental change in the United States, as is well known. By comparison, the occupational healthcare system is less well known and much less well understood in its complexity. The occupational healthcare system is separately financed and is organized around different principles than the general healthcare system. The two systems share practitioners and facilities, but they function very differently in the service of each system and often interact poorly.

The currently decentralized and individualized approach to providing medical care has fragmented care and resulted in a diffuse and sometimes incomplete network of practitioners providing OHSs. These practitioners primarily serve a centralized workers' compensation system and coexist with medical services sponsored by larger employers. Large parts of the system, such as the appeals process in workers' compensation and the in-plant occupational health and safety activities, have become invisible to the average practitioner in the community.

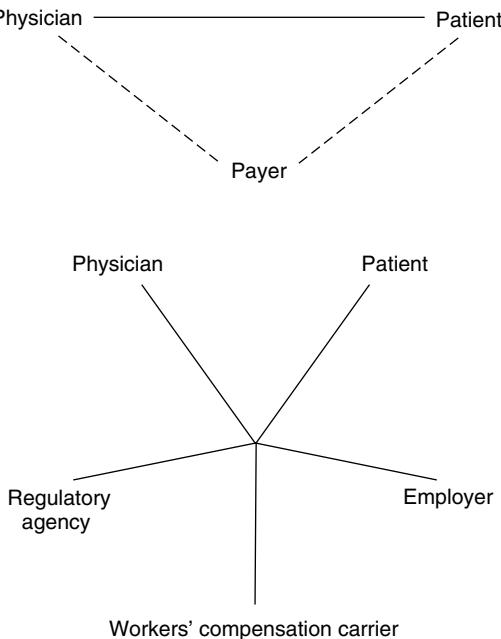
OHSs are often the historical basis for general systems of health-care. In many countries, and in North America in times past, OHSs have been the basis for providing general healthcare in newly urbanized or settled areas of industrial development. The railroads played a major role in establishing hospitals and supporting physicians outside the major cities in both the United States and Canada in the eighteenth century, for example. In Venezuela, health services for the general population in the western part of the country developed from medical services provided for the employees of oil companies. In Mexico, the financing of individual health services started through the employment-based social security (*seguridad social*) systems. In the United States, innovations in healthcare financing for employees and their dependents laid the foundation for health maintenance organizations, capitated reimbursement, and integrated hospital systems.

### Characteristics of the System

Occupational healthcare as provided by the physician is part of a much larger, complex system. The system is built on different assumptions than the general or personal healthcare system. In a traditional fee-for-service setting, the primary relationship is that between physician and patient, with the insurance carrier or other third-party payers now playing an increasingly active role, as illustrated in Figure 20.1. OHSs, on the other hand, are often five-sided relationships involving the physician, the patient, the employer, a government regulatory agency, the workers' compensation board and



**Figure 20.1.** Relationship of the physician, patient, and other stakeholders in general medical care.



**Figure 20.2.** Relationship of the physician, injured worker and other stakeholders in occupational medical care.

carrier, and often a union as well, as illustrated in Figure 20.2. Among these stakeholders, information circulates subject to legal requirements and accepted rules of confidentiality. Authority for decisions and responsibility for compensation do not rest mostly or even primarily with the physician but are shared according to the roles of each player. The physician often acts as an agent reporting to the employer or to the carrier rather than on behalf of the individual patient, as in the case of independent medical evaluations. Figure 20.2 illustrates these differences. Ethical dilemmas predictably arise when these responsibilities conflict, but the legal relationships are fairly well defined.

Occupational medicine services may be divided into two categories: (1) curative and ameliorative and (2) preventive (Table 20.1). Curative and ameliorative services are intended to cure or to limit disease or to manage an existing problem. Preventive services seek to

**Table 20.1.** Occupational Medicine Services

Subsets of Services	Curative/Ameliorative	Preventive
“Industrial medicine” (traditional name, not a recognized specialty)	Acute and chronic care Disability evaluation	Pre-placement and periodic screening
Occupational hygiene (teamwork required)	Health hazard evaluation Compliance with government regulations	Consultation Worker education Surveillance Monitoring
Personal health (individual health)	Employee assistance programs	Health promotion (enhancement)

avoid exposure of the worker to hazards, to detect disorders at an early and potentially curable stage, and to limit disability. The physician's responsibility is first to the patient whether in treating occupational disorders or intervening and educating to avoid the development of preventable illness. It is the employer's role to deal with the economic and public relations aspects of a problem, not the physician's role. Table 20.2 provides a breakdown of the tasks performed by a physician-led regional medical office of one large utility in the

**Table 20.2.** Distribution of Services Provided by or Under the Supervision of a Physician for a Large Utility in a Representative Year

Management of work-related injury or illness, clinical	30 percent
Fitness-to-work evaluations, other than pre-placement	25 percent
Absence review, certifying return to work after illness, and reviewing personal healthcare issues	20 percent
Counseling and arranging referral for employee assistance	8 percent
Periodic health evaluations, either surveillance mandated by regulation, or voluntarily for employees at elevated risk	8 percent
Management of minor personal health problems, clinical	6 percent
Pre-placement evaluations	3 percent

United States. It illustrates the distribution of work in an industry without unusual hazards.

In occupational medicine, the diagnosis is often the beginning rather than the end of the evaluative process. This is particularly true for occupational illnesses due to workplace exposures and for repetitive motion injuries. The related but distinct questions of the exposure responsible, the work-relatedness of the condition, and the expected degree of disability are sometimes more important than a precise diagnosis because on such questions hinges eligibility for compensation, prognosis for rehabilitation, fitness for future work, and prevention for the protection of other workers.

Prompt and effective treatment of the injured worker can save that individual from unnecessary disability. An early return to work, when medically indicated, can also assist in rehabilitation. Acute care in an occupational healthcare setting, however, is more than a simple service function as it might be in a hospital emergency room. Opportunities exist within a well-organized occupational healthcare system to use the lessons learned from each injury, either singly or as aggregate statistical data, to prevent future injuries. The ultimate goal of an occupational health intervention, of course, is to control the hazard that led to the problem in the first place.

The physician never acts alone in an occupation-related case. Each action is reviewed and closely monitored behind the scenes, whether the physician is aware of it or not. Workers' compensation carriers require telephone authorizations after the initial visit, copies of the medical record, and supplemental reports before payment is authorized.

Another aspect of delivering occupational healthcare is the importance of accurate and segregated medical records. These records are subject to review not only by the patient, but also by the insurance carrier, the employer, or outside consultants when a claim is appealed. In disputed cases, medical records are subpoenaed. Occupational health records must be maintained separately from the patient's personal medical records to prevent inadvertently breaking confidentiality. The occupational health record should not include personal or medical information that is not directly pertinent to the workers'

occupational health, such as incidental past medical history (which may include much confidential material of no concern to other parties), detailed family history, correspondence about billing, or any remarks on the patient's character or that of the employer. The rules for disclosure and for authorizing release of information by the patient are spelled out in occupational health and safety regulations. Occupational health records should be maintained for at least thirty years (in keeping with the OSHA standard for retention of medical records) and preferably indefinitely.

A clear understanding of the occupational health and safety and the workers' compensation system will prevent unnecessary problems and administrative burdens. Information on how each system operates is available on request from the appropriate federal, state, or provincial agency, workers' compensation board, or from workers' compensation insurance carriers. Information on maintaining occupational medicine records is available from the government agencies responsible for occupational health and safety.

## **OCCUPATIONAL HEALTH PROFESSIONALS**

Teamwork, administrative ability, and versatility play a greater role than in other specialties. The new occupational and environmental medicine (OEM) physician soon finds that he or she depends on other skilled health professionals, such as occupational health nurses, safety engineers, and industrial hygienists (engineers trained in the recognition and control of health hazards in the workplace). Each of these professional groups has its own training, certification, and licensing arrangements; the OEM physician relies as heavily on these skilled professionals as a surgeon relies on the operating room team. Some of the professions are listed in Table 20.3. For ease of communication, it is useful to know about the background and role of the principal occupational health professions.

There are two types of OEM physicians traditionally found in employer-based facilities: the "plant physician" and the "corporate medical director." Since the mid-1980s, most OEM physicians are

**Table 20.3.** Occupational Health Professions

---

Occupational Medicine
Occupational Health Nursing
Industrial Hygiene
Ergonomics
Safety Engineering
Radiation Health
Occupational Audiology
Toxicology
Epidemiology
Risk and Liability Control
Employee Assistance Management
Vocational Rehabilitation
Fitness and Health Promotion Program Management

---

based in the community and practice out of their own medical offices or other OHSs; they are formally called “community-based occupational medicine practitioners” (and often informally call themselves “occ docs,” pronounced “ock-dox”). “On-site medical practitioners” may be as good a name as any for the increasing number of medical practitioners who practice in “on-site clinics,” general medical clinics in stores and factories. This trend only began in 2006 but has grown rapidly. On-site practitioners are not OEM physicians, but they are sometimes called upon to provide basic services, such as pre-placement evaluations.

“In-plant services” are provided inside the employer’s facilities, usually by a practitioner hired part-time by the employer. A “plant dispensary” or infirmary may employ a plant physician and an occupational health nurse whose responsibility may range from simple triage and first aid to fairly comprehensive primary care. Plant physicians have no direct authority in setting corporate policy. The corporate medical director, on the other hand, is usually based at the headquarters of the organization.

Corporate medical directors usually see only a few selected patients with special problems and spend most of their time managing

and contributing to policy on health-related matters within the company.

Community-based occupational medicine practitioners became a majority in the 1990s, reflecting outsourcing and downsizing in corporate medical departments. Increasing numbers of trained OEM physicians started practicing outside industry in clinics, group practices, and other community-based medical facilities. Paradoxically, the total volume of occupational medical services seems to have dramatically increased at the time, probably because the new availability of such services in the community attracted midsize and smaller employers that previously did not access OHSs. By 2000, the total number of practitioners had increased disproportionately to the number of physicians leaving industry and the number of specialists finishing training programs, indicating that more practitioners were recruited into the field in midcareer. A small number of physicians, mostly specialists in occupational medicine, maintain full-time private practices exclusively directed to disability evaluations, assessing causation and work-relationship, and reviewing cases under appeal or litigation.

On-site medical practitioners are mostly family physicians, who practice primary care rather than the full range of OHSs. To date, these clinics have not employed OEM physicians. However, the convenience and low cost of the model means that inevitably the practitioner will be asked to perform many services in OEM.

An “occupational health nurse” specializes in OHSs, almost always in an in-plant setting. Occupational health nurses are trained in health education, occupational health, disease recognition, rehabilitation, and administration. The specialty is highly regarded within nursing for its relative autonomy and level of responsibility. “Occupational health nurse practitioners” perform midlevel practitioner duties in the workplace, much like plant physicians. The principal organization for nurses in the field is the American Association of Occupational Health Nurses. Since 1972, there has been a specialty board certification in North America for nurses in the field.

An “occupational hygienist” or “industrial hygienist” is a professional specialized in the recognition, evaluation, and control of occupational health hazards. The training includes engineering solutions, ventilation, analytical chemistry, mathematics, and toxicology, usually at the master’s degree level. Board certification is provided by the American Board of Industrial Hygiene and is quite rigorous. The principal organization for industrial hygienists in both the United States and Canada is the American Industrial Hygiene Association, which has many local chapters.

A “safety engineer” is responsible for the recognition and control of safety hazards. Their preparation is usually not at the graduate university level. Many safety professionals have obtained their training in short-term, intensive institutes or seminars. Often, safety professionals have worked their way up through experience and continuing education from positions as foremen or supervisors. Their training usually emphasizes physical hazards, fire protection, and safety education. In the United States, the American Society of Safety Engineers and the National Safety Council are the leading professional and general membership organizations concerned with safety. The Canadian Society of Safety Engineers and the Canada Safety Council are the major general safety organizations in Canada. There are many associations concerned with safety in particular industries and many more committees and working groups within trade association and industry groups.

A “work evaluation” or “rehabilitation counselor” is a consultant called into a case by the workers’ compensation or other insurance carrier to assess the work skills, physical tolerances, specialized training, and motivation of the worker. The assessment is used in judging disability in workers’ compensation cases. Work evaluation counselors are usually trained at a bachelor’s or master’s level and often have backgrounds in rehabilitation and physical therapy as well.

An “audiologist” is a health professional specializing in the evaluation of hearing disorders. Audiologists are usually trained at the master’s or doctoral level in behavioral audiology, the physiology of hearing, speech pathology, acoustics, and psychological aspects of hear-

ing impairment. Audiologists usually work closely with physicians in the diagnosis and evaluation of hearing impairment. Audiologists are not to be confused with occupational hearing conservationists, who are technicians with a certification from the Council of Accreditation in Occupational Hearing Conservation, which attests to their skill in performing audiometric evaluations of workers exposed to noise. Occupational hearing conservationists are technicians trained in the proper procedures for conducting screening, but not necessarily diagnostic, audiograms in working populations.

## **OCCUPATIONAL HEALTH SERVICE FACILITIES**

The role of the OHS is to provide assistance both to management and the employee/worker so that the worker is not harmed in the course of work. To fulfill this role, the OHS facility must be and also must be perceived to be objective and impartial, whether it is part of the management structure or an external healthcare provider. The operational responsibility for sound occupational health and safety practice lies with the people who do and supervise the work: the employee, operating personnel, and managers. The OHS exists to serve them; it can assist, but it cannot be in all places at all times. Its role is that of consultant, troubleshooter, teacher, auditor, and advocate. It cannot function as a policeman, personnel officer, or scapegoat and remain effective.

Most routine occupational health services are simple in their execution. Employers are not always aware of differences in quality among potential providers of occupational health services and may not see any reason to pay a premium for quality when “adequate” will do, insofar as acute care for simple injuries is concerned. Perceived quality of care, beyond adequate, is not usually a consideration. Although the worker receives occupational healthcare in the role of “patient,” it is usually the employer (sometimes the carrier) who selects the provider of care and plays the role of “consumer.” In this situation, the major incentive governing the behavior of the “consumer” is to control costs, not to maximize benefit.

The decision by the management of an employer to develop an in-house capability to provide or to contract with an outside medical provider for an OHS and how the arrangement shall be structured must reflect the realities within both the organization and the community of which the organization is a part. Among the considerations that managers should take into consideration are the following:

1. Is qualified medical care (i.e., from physicians knowledgeable about occupational disorders) available in the community and easily accessible?
2. Do operations within the organization require special insight or expertise because of particular hazards, unusual technology, or the potential for injuries?
3. What are the pertinent regulations regarding occupational safety and health?
4. What has been the past experience of the organization and what is its expected future growth?
5. Who are the employees, what is their social, ethnic, and economic background, will they use the facility, and can the facility effectively meet their needs in communication and for prevention and health promotion?
6. How much of an investment is the organization prepared to make?

### **Types of Occupational Health Facilities**

Occupational health facilities can be divided into “in-plant” and “external” facilities. Until recently, in-plant services dominated the delivery of occupational healthcare.

In-plant services are provided within the employer’s facilities and almost always by a practitioner employed by the employer. In-plant services are practical only for larger firms. Within the last two decades, “off-site” or community-based healthcare facilities have

displaced in-plant services as the principal form of organization for providing healthcare. Many practitioners, especially family physicians, provide such services in the course of their practice out of their own office. The major facility of this type at present, however, is the free-standing “occupational medical clinic,” a medical group or practice more or less devoted to providing healthcare on a contract or fee-for-service basis with local industry. (In past years, the term “industrial medicine clinic” was often used, implying an inferior level of care.) Large multispecialty group practices, particularly in the east and southwest, have been developing occupational medicine services as a major new thrust. Hospital-based OHSs are becoming more prominent. At one time, there were many union-sponsored occupational health centers, but now only a handful of labor-oriented occupational health clinics exist, principally in New York and Canada, most notably the provincially funded Occupational Health Clinics for Ontario Workers (OHCAW). Government agencies may have medical clinics for their own employees. The military maintains occupational health clinics for both active duty and civilian personnel and in the aggregate constitutes a major subset of the occupational healthcare system.

OHSs in industry can be divided into “in-plant” or internal and “off-site” or external facilities and services. Internal services are overseen by a plant physician or occupational health nurse. On the corporate level, discussed later, a corporate medical director may oversee the in-plant services or coordinate and supervise the delivery of care by off-site contract physicians, or a combination of both.

Occupational medical services may take one of four basic forms:

- Internal, serving one employer
- External or “community-based,” serving one employer or, more often, one industry
- Internal, serving multiple employers
- External or “community-based,” serving multiple employers

## **Internal, Single Employer**

Internal OHSs serving one employer are typical of large corporations, large single plants, and public agencies of cities and higher levels of government. They exist on two levels: the in-plant health service and the corporate medical department. In the past, the in-plant health service usually went by the name of infirmary, dispensary, or clinic; today it is usually called the occupational health or employee service. It is a facility where medical care above basic first aid can be rendered to ill or injured employees and where programs of a medical nature for prevention and health promotion are organized and implemented. It may be supervised by a suitably trained nurse or by a physician and is often staffed by both with the physician available by contract at designated times during the week. The in-plant health service is the first level of care, providing initial treatment and evaluation for fitness to work but rarely managing complex or difficult cases, which are usually sent outside to local practitioners. Unless a plant presents an unusual or particularly serious hazard, such as a shipyard, or it is in an unusually isolated location, it is generally unwise for the in-plant health service to provide specialized or major surgical care because of legal liability and the cost of maintaining adequately equipped facilities.

On-site general medical clinics represent a rapidly emerging model, newly introduced in 2006, although similar arrangements existed in the early twentieth century. Some companies, led by Wal-Mart, have opened general medical clinics in their stores and factories. The objective is to provide additional services to customers that add to convenience, increase sales, and make the store indispensable to the community. Wal-Mart began with plans for 400 such clinics and contracted with providers to staff them. At least one, Checkups, subsequently closed twenty-three facilities in 2008, but this contraction appears to have been due to inadequate financing rather than because business model was unsuccessful. Major drug store chains have opened similar clinics in their stores, including CVS and Walgreen's, with 200 clinics. Major employers such as Disney,

Toyota, and Harrah's Entertainment have opened clinics in their plants and office facilities as well but not to attract customers or visitors. These clinics operate with the objective of reducing health-care costs, time lost due to scheduled and emergency medical care, sickness absence, and presenteeism (when an employee is at work but not functioning due to a health problem). The on-site medical practitioner is not an OEM physician but is often called upon to perform basic procedures such as injury care and pre-placement evaluations. At least one contractor, CHD Meridian, offers a range of services including fitness-for-duty evaluations, pre-placement evaluations, employee assistance, acute injury care, impairment evaluation, and some health education programs; and markets its ability to handle workers' compensation cases. This model represents a potential opportunity for OEM physicians who wish to provide primary care as well and will probably do well in the current recession because of low cost and convenience.

### **Community-Based, Single Employer**

External services serving a single employer or a single industry are common in the form of individual physicians or consultants who do part-time work, usually on an as-needed basis, for a particular company or public agency or a narrow range of companies in the same economic sector. Usually, this activity supplements a private practice, but it can be part of a physician's retirement activities or the result of a special relationship or prior experience with the agency or company or a specialized industry, such as aviation, with particular needs.

### **Internal, Multiple Employers**

Internal programs providing services to multiple employers are of three general types: (1) in-plant services open to subsidiaries and affiliates of a larger company (essentially an extended form of the internal-single employer approach), (2) in-plant services open on a

contract or fee-for-service basis to other small local companies (a variant of the external, multiple-employer approach), and (3) contractual arrangements by which an OHS provides limited in-plant services for several employers at the same time. The first two are more common in other countries where the healthcare system is not well developed or where “polyclinics” (healthcare centers serving the needs of a particular population) were once organized by industry or employment, as in Eastern Europe. The most common situation of contractual arrangement serving multiple employers with in-plant services is when a physician or occupational health nurse works part-time in more than one plant, practicing in the plant facilities.

### **Community-Based, Multiple Employer**

External services serving multiple employers predominate in the occupational healthcare system. They include individual physicians, group practices, hospitals, and clinics that accept patients from several local employers on a contract or fee-for-service basis. OHSs serving many employers are in a relatively poor position to influence management and are at a disadvantage (and are not always motivated) in providing services in prevention or control of hazards. This is probably the only practical model for providing services to small enterprises, however, given the high overhead costs of maintaining internal services for small numbers of employees.

Several models exist for the provision of occupational healthcare to many employers. The most common are:

- Occupational medicine in primary care practice
- Consultation practice in occupational medicine
- Multispecialty group practice
- Hospital-based clinic
- The occupational health center (successor to the “industrial medicine” clinic)

The delivery of occupational health services at a high standard requires administrative skills beyond office management. To communicate successfully with all parties, including workers' compensation boards, employers, workers, regulatory agencies, and insurance carriers, requires special skills and the ability to appreciate a problem from many points of view. Clinical occupational medicine practice often requires the screening of large groups of well persons, in situations in which price is a factor. The various models of occupational health-care facilities do not do equally well on these requirements.

Independent OHSs, as combined with primary care, as a consultation practice, or freestanding occupational health center, are entrepreneurial activities in which the business plan centers on the success of the OEM practice, which depends on the practitioner for quality of care and delivery of service. Sponsored OHSs, which are part of a hospital or group practice, are business-driven entrepreneurial activities in which the service is expected to fit in with a broader business plan and depends on the practitioner for quality of medical care but on the institution for delivery of service and support.

OHSs that are owned, operated, or embedded in healthcare institutions, such as multispecialty group practices and hospitals, face a common set of problems that is discussed in the subsection on operations.

### ***Occupational Medicine in Primary Care Practice***

Individual practitioners and hospital emergency rooms deal with innumerable occupation-related cases as a matter of routine, usually providing only acute or episodic care. This diffuse network of practitioners is the most widespread network of providers, but its capabilities are limited by the lack of available technical expertise; the absence of well-exercised skills in aspects of care management unusual or unique to occupational medicine practice; and the constraints on time, energy, and resources that can be invested in a narrow aspect of private practice. Most primary care physicians who do occupational medicine "on the side" are family physicians or emergency medicine specialists, although some are internists.

Primary care physicians have a decisive role to play in the private practice of occupational medicine. There are nowhere near enough OEM physicians to see all occupational injuries and illnesses. Primary care physicians will provide the major share of OHSs for the foreseeable future—and probably well beyond. Physicians may choose to emphasize occupational medicine in their practice for a number of reasons:

- To gain an additional source of revenue
- To obtain an additional source of patients for general care
- To meet the need for services in the community
- To conduct clinical research or to provide education and training
- To pursue a personal interest

There are inevitably practical problems for the primary care physician working in OHSs. These include the large administrative burden that falls on the physician because of the need to prepare claims and numerous reports, frequent legal actions involving the physician as a witness, and the need for frequent reporting to employers. On the other hand, the concept of prevention is directly compatible with the missions of family practice and internal medicine, in particular.

The greatest drawback to the incorporation of occupational medicine practice into primary care is the patient-centered role one plays in primary care versus the client-enterprise-provider relationship of occupational medicine. There is also a distinct difference in culture among different specialties. The cultural dimension goes deeper than a lack of understanding among primary care physicians about occupational health services, although that can be an obstacle. The world of occupational medicine is quite different from that of most other specialties and requires a detached view of the injured worker's problems that differs from the patient advocacy role in primary care. The practice of occupational medicine can be very disconcerting to physicians oriented to traditional primary care.

Family practitioners are in an excellent position to provide the wide range of clinical services, including minor injury care, that may be needed. The family physician is trained and acculturated to be the advocate of the patient, to provide a high level of basic care, to establish long-term relationships and continuity of care, and to know when to refer patients to a specialist for clinical care. This role may conflict with the need in occupational medicine to report findings that may not appear to be in the patient's immediate interest (such as a lower rather than a higher level of impairment), with the extensive documentation needs that are essential to occupational medicine (especially in workers' compensation), to separate the patient's satisfaction with objective criteria in making decisions about fitness to return to work, and to know when a work-related problem requires investigation and consultation with professionals outside the world of medical specialties, such as occupational hygienists.

The internist brings strength in analysis and deductive reasoning that is particularly important in occupational disease cases. Internists are trained and acculturated to be master diagnosticians, but the diagnosis itself is often less important in occupational medicine than causation and impairment. Internal medicine training is often weak on musculoskeletal disorders and toxicology and may not provide some of the relevant clinical skills.

Emergency physicians are trained and acculturated to provide excellent care in urgent situations and have diverse clinical skills, but the care they give is episodic. Physicians with such training need to be prepared for the other side of occupational medicine practice, the preventive services, and the need for tracking recovery over weeks and months, in addition to monitoring people who have nothing much wrong with them. The few emergency physicians with training in toxicology are particularly well equipped to deal with occupational diseases.

The occupational healthcare system is an often complex network functioning somewhat isolated from the mainstream personal healthcare system and subject to its own internal tensions and forces.

One of the issues OEM physicians continually face is the assumption that they perform mainly acute injury care that can be handled equally well by primary care practitioners. Occupational medicine practice, particularly at the higher levels, tends to be less clinical and more administrative and cognitive in the skills it demands, requiring a deeper knowledge of population health, toxicology, fitness for duty, occupational hazards, functional assessment and the workers' compensation system. The OEM physician must deal continually with strong economic, political, and social pressures even above the present complexities of medical practice. These include workers' compensation, disability claims, lawsuits, government regulations, labor-management relations, budgeting, and other issues far from usual office practice.

Any physician moving into OEM practice from a primary care specialty is welcome but is well advised to take the time to prepare by mastering essential skills in OEM. Some physicians simply choose not to make the investment and therefore do not practice at a high level.

### ***Consultation Practice in Occupational Medicine***

A small number of full-time private practitioners in OEM, who are usually located in major cities, practice primarily as consultants. Their practice may be associated with academic institutions or independent, whether or not they hold an academic appointment.

In the setting of academic institutions and teaching hospitals, occupational health consultation clinics combine clinical practice with teaching and research. Most patients to such clinics come by referral from other physicians. They mostly serve as independent medical examiners in workers' compensation cases, as expert witnesses in disputed cases, and as consultants in toxicology, epidemiology, and often aerospace medicine. (Many aerospace medicine specialists also practice general occupational medicine.) The organization representing such clinics is the Association of Occupational and Environmental Clinics.

Consultation practice in occupational medicine today increasingly requires meaningful credentials and several assured sources of referrals. A referral network normally comes from visibility, reputation, success in managing certain types of cases, and extensive contacts with employers. If one wishes to develop a referral practice, one must ask where these cases are to come from, who will see them first, and why they should be referred to a particular consultant over another.

Occupational medicine is practiced in a fishbowl. Opinions and findings are under review by workers' compensation boards, employers, insurance carriers, and frequently unions. Each is a potential source of or a potential obstacle to referrals. They are very sensitive to inadequate documentation, incomplete records, and delayed reports and react silently by steering referrals elsewhere.

Complex cases, those which are most likely to be referred to a consultant, often go to court or to arbitration, and the physician must be prepared and be seen to be prepared to stand behind opinions and reports. The suitability of a physician as a possible expert witness when the possibility of litigation exists is often a factor in the selection of a consultant in a given case. The physicians' formal credentials in occupational medicine become important in such situations. (See Chapter 23.)

Occupational medicine consultation clinics, unlike primary care clinics for occupationally related injuries, are seldom profitable in the private sector unless overhead is kept to a minimum and referrals include many "high-value" cases (high value in this sense only in terms of financial worth, of course), which may involve medicolegal support, independent medical evaluation, or high-value services such as exercise testing and cardiopulmonary evaluation. One way to keep costs low is to rely instead on tests performed by lower-paid staff and limit the patient-physician interaction to situations where the physician is most needed: a highly focused interview and examination. One way to encourage a steady flow of high-value cases is to associate with one or more OHSs that provide acute injury care and a primary care level of healthcare to injured

workers. Acute-care facilities generally prefer not to deal with complicated cases or cases of occupational disease and are happy to refer them in order to get them out of their system, which is geared to high-volume care.

### ***Multispecialty Group Practices***

Multispecialty group practices are partnership organizations in which specialists in compatible areas of medical practice engage in a relationship in which common expenses and resources are shared, whether in a single large clinic or a network including satellite clinics. Group practices are often faced with intense competition and feel the need to capture groups of patients, as in health maintenance organizations and similar prepaid plans. Group practices in the United States have often entered OHSs to "lock-in" large groups of employees that are then expected to use the OHS services for their personal health needs and those of their families. Historically, multi-specialty group practices have done well in providing occupational health services, and most of the larger practices have at least some physicians involved on a routine basis.

Multispecialty group practices do have certain disadvantages. They have usually developed as providers of personal healthcare and often developed occupational healthcare services without appropriate preparation. Group practices tend to enter the field by either creating within themselves new services staffed by physicians recruited for the purpose or by using the personnel they have on staff already, qualified or not. As the OHS develops, it frequently finds itself subordinated to personal health services whenever decisions are made regarding allocating resources, recruiting staff, marketing, and opening satellites. For this reason it is advisable to keep the administrative structure of the OHS as autonomous as possible within the group, even at the expense of some duplication. Administrators who are not sensitive to employer's concern over lost time may see no need to expedite registration and waiting periods. One of the major criticisms of group medical facilities by employers is that because

occupational injuries are no more important to the group than nonoccupational injuries, an injured employee may have to wait longer to be treated because the injured employee will be competing with nonoccupational injuries for available physician time. Groups suffering financial problems or poor cash flow may also be tempted to shift overhead expenses onto the apparently lucrative occupational services contracts, raising costs to employers and destroying the attractiveness of the group-based option.

### ***Hospital-Based Occupational Health Clinics***

Hospital-based OHSs have been introduced in many large institutions in the United States. More recently, community hospitals have entered the field in large numbers. Most hospitals have promoted their services through selective business contacts rather than media advertising. Like group practices, the incentive for hospitals is to cultivate a patient base likely to enroll in their health maintenance organizations and to use the hospital and its outpatient clinics for their personnel healthcare. Some hospitals do a very credible job providing specific occupational health services, but others develop their programs as a marketing device, with little real commitment of resources or quality assurance.

Most of the services are relatively rudimentary, operating either as an extension of emergency room care or an isolated community outreach program. A few, however, are well developed and comprehensive in their provision of services. Pacific Presbyterian Medical Center in San Francisco has provided an extensive service for many years. Many academically oriented medical centers that have offered such services in the past did so as part of training programs in occupational medicine. Many hospitals are using occupational health services to build up “wellness,” or health promotion, programs, and virtually all expect the services to increase utilization of other hospital departments such as radiology, clinical specialties, and laboratories. Unfortunately, the fees charged by hospitals for routine services such as chest films are often noncompetitive in the marketplace.

### *Occupational Health Centers or Clinics*

“Occupational health centers” or clinics are free-standing facilities, serving many employers in a well-defined geographical area such as a town or an industrial park. Although not a new phenomenon, these clinics have become attractive models for entrepreneurs interested in providing direct healthcare services in settings with low overhead. Their patients are generally drawn from local employers, with whom they have contracts or informal arrangements to provide services, emphasizing episodic care of injuries or illnesses on an acute basis and periodic health evaluations as mandated by law or by company policy. The occupational health center is characterized by certain basic features: a central facility serving multiple client employers, expansion of medical staff with primary care providers, an emphasis on medical care with less involvement in prevention, close attention to market trends, and a relatively small nursing and support staff. Their revenues are generated primarily from workers’ compensation fees for the former and third-party payments from the employer for the latter. In a few cases, they have grown quite large and diverse. In general, however, they grow by establishing satellites or spin-off clinics and by creating larger healthcare systems composed of several clinics.

Occupational health centers focus on cost-effective management of minor trauma and pre-placement evaluations. The business model emphasizes high-volume care, triage for cases that are complicated, and high efficiency of operations to keep overhead under control. The primary advantage to injured workers and employers is rapid access to care, early return to work, one-stop provision of all essential services, and correct and timely reporting and filing of paperwork, especially that required by workers’ compensation and the U.S. Department of Transportation for commercial drivers (which is very time sensitive).

Occupational health centers are a new generation of service that grew out of the “industrial medicine clinics” of an earlier era, peaking in the 1970s and early 1980s. At that time, industrial medical

clinics generally had a reputation for marginal quality of care and isolation from the mainstream of occupational medicine. One exception was the Detroit Industrial Clinic, which was a model of its type. During the outsourcing of the 1980s, many well-qualified OEM physicians left corporate medical departments and entered practice in the community, and other physicians entered occupational medicine practice because of the new opportunities. Employers that had previously been served by in-house health services were more demanding in their expectations. The result was a substantial upgrading in that sector of healthcare.

The proliferation of occupational health centers and the emergence of these facilities out of the old industrial medicine clinics have led to many favorable changes in OHSs. It has made services that were previously available only to major employers easily accessible to small enterprises. The sector has consolidated in recent years and is now heavily dominated by a few providers, the largest of which is Concentra.

## **CORPORATE OCCUPATIONAL HEALTH SERVICES**

Large employers, by definition, have more than 1,000 employees and often need a full-time, in-house corporate medical department. Sometimes this corporate service acts primarily to coordinate a network of individual healthcare providers at scattered workplaces, either plant physicians or community physicians receiving employees at their own facilities. In other organizations, the corporate medical department primarily manages contract services and personnel that perform OHSs as needed. After the downsizing and delayering of the 1980s, many large corporations outsourced most of their occupational health and safety services, including medicine, to external providers. Often, these providers were their previous employees. The result was that there are many fewer corporate medical departments today than twenty or thirty years ago, but the total number of providers is greater. The corporate medical departments that have survived often assumed a broader span of responsibility, including

occupational health, health affairs management (such as insurance and managed care issues), safety, and the environment. Recently, many corporate medical departments were beginning to hire again, in small numbers, to bring OEM physicians back into the organization to assume responsibility for this expanded portfolio and to manage the fragmented contractor relationships that developed after their OHSs were outsourced.

### **Corporate Medical Department**

The corporate medical department is a unit within the management structure that reports at a high level and participates in policy formation and decision making. The corporate medical department is headed by a physician, usually with an executive title and position, who directs and reviews the provision of healthcare within the organization, including its plants and other worksites, the headquarters staff, and the medical aspects of the benefits packaged for employees. At the corporate level, direct medical services may be provided to local or headquarters staff or in the evaluation of individual problem cases, but this function is secondary to that of co-ordinating and advising on health-related matters within the organization.

The single most important determinant of cooperation in matters regarding occupational health and safety is a clear and explicit corporate policy statement, widely distributed, reinforced with regular application, and referred to often in business communications. This clear commitment should be reflected in personnel policies and budgeting and cited throughout the corporate policy manual. Among its other elements, the policy should hold all personnel accountable (including top managers, supervisors, and workers on the shop floor), commit the employer to meet or exceed applicable government regulatory requirements, hold contractors on-site equally accountable for unsafe practices by their own employees, and commit management to regular review and reporting of their performance in health and safety.

The physician should report at the highest levels of management, preferably to the chief executive officer or at the vice-presidential level, to be maximally effective. This ensures that issues in occupational health and safety are reported without distortion or censorship and permits the corporate medical director to convey a set of priorities that may not be shared by midlevel managers. It sends a signal throughout the rest of the organization that occupational health will not be subordinated to other priorities.

Occupational health and safety deals with working conditions and hazards and other issues of the quality of life for workers. Ill or injured workers contribute to absence from the job, and their paperwork is usually handled by the personnel office. Many employers have therefore placed their occupational health and safety units within the personnel or "human resources" department. The advantage of this system is that it appears to be tidy administratively and facilitates the handling of absence data and claims for workers' compensation. The disadvantage is that it undermines the integrity of the occupational health and safety unit. Placing it so close to the human resources offices and functions compromises the perceived and, usually, actual independence of the occupational health and safety unit. It also tends to push issues involving absence onto the OHS, when this is properly a human resources function, not a medical function. It becomes too easy for HR managers to request (and insist upon) access to information to which they are not entitled. Workers seeking help or involved in exposure-related issues draw the conclusion that the HR manager is watching their actions and that they will be laid off for complaining or for becoming injured or ill. Often, this concern is justified because the personnel officer does not understand the proper limits of his or her authority. Physicians practicing in such an arrangement are severely compromised, appearing to be too close to management to be trusted. The inevitable result is withholding of information, refusal to cooperate, reluctance to accept medical judgments as authoritative, and suspicion of any new health-related projects.

Even worse is placement of the occupational health and safety unit in a department of labor relations. Preoccupied with issues of union

negotiation, managers in these departments often tend to withhold or censure information that should be shared freely with workers. The occupational health and safety unit is often dragged into the middle of labor-management disputes and forced to declare itself on the side of management. This forced display of organizational solidarity may please management, but it inevitably destroys the credibility of the occupational health and safety unit and turns the service and all associated healthcare professionals into adversaries of the worker. Without confidence in the unit, workers will not only withhold information and fail to cooperate, but many actively resist programs that might benefit them because they are seen as being imposed by management and subject to a hidden agenda.

Although managers understandably wish for all senior employees in their organization to be on the same team, close identification of occupational health and safety personnel with line management is the kiss of death for their ability to work closely with workers. Since this ability is key to their effectiveness, the employer is much better off accepting that the OEM physician, nurse, or hygienist is a professional within the organization who must function somewhat independently and must avoid an adversarial role with workers. At the same time, the OEM physician, nurse, or hygienist must be capable of understanding management systems and objectives and communicating with management in language that can be understood.

A successful corporate OHS is one that is appropriate to the needs of the organization it serves, efficiently managed, ethical and respectful in treatment of the worker, and effective in relation to cost. However, in order to survive and be successful, the corporate department must constantly justify its existence to senior management, especially financial officers of the company. Corporate OHSs are very vulnerable as business units because they do not directly produce goods and services, they are perceived as “consumptive” cost centers (units that cost money but add little value), and are hard to evaluate, especially if they are successful.

All large organizations have a common basic structure. The “line of authority” extends from the top leader to the employee at the

lowest level. Along this line all management decisions and orders are conveyed relating to the goods or services that are the business of the organization. The staff line, perpendicular to the line of authority, supports it by providing necessary services such as financial, legal, human resources, and occupational health and safety. Occupational health and safety is intrinsically a staff rather than a line management function, but it must necessarily be a line responsibility if safe work practices are to be followed consistently. The occupational health and safety staff cannot be all places at all times and cannot know the workplace as well as the workers who are assigned there regularly. If staff bear the primary corporate responsibility for occupational safety and health, the line managers may not be as responsive or interested and hazards will not be corrected. Individual managers and supervisors must be individually responsible for the safety performance or injury experience in their areas or responsibility and should be individually evaluated for performance in this area just as they are on productivity.

On the organization chart, the corporate medical director is in a "staff" position, providing support services and assistance to top management, rather than in a "line" position of direct authority over production and services. Those occupational health personnel under the direction of the corporate medical director should cultivate close relationships with line personnel, however, in order to ensure cooperation and full understanding of the production process and options for control of potential hazards.

An important role of the OEM physician in a corporate medical department is to manage contract services, such as audiology and hearing conservation. A successful service is rarely found "off the shelf." Obtaining the right system to meet the needs of the employer requires careful analysis and good planning. The variety of "packaged" occupational health programs for sale in the marketplace does not obviate the need for planning and evaluation because many of these package deals may be of little or no value to the employer and may not provide a service of acceptable quality for the worker. They may be poorly conceived, badly implemented, inflexible, or unsuitable

to the employer's particular situation. Lacking a plan and carefully considered expectations, an employer can easily fall into the trap of paying for a system or program that fails to meet the real need.

The department should have a close working relationship with the safety and occupational hygiene functions. Safety and hygiene usually report to an operating manager, possibly with other responsibilities for risk and loss control. Once a company has grown to around 3,000 employees, a single administrative grouping of health, safety, and hygiene should usually be formed under a senior manager, preferably the medical director if the medical director has management skills for the responsibility.

The current trend is to consolidate health, safety, and environmental affairs into an overall "HSE" department. HSE departments combine health, safety, and environmental management, often under the leadership of a manager specializing in the field. In such arrangements, the OHS may be a unit within the HSE department, the non-medical OHS and environmental functions may be integrated into the management structure of the OHS, or the physicians in the organization may be part of a team addressing HSE issues with additional clinical responsibilities. A few organizations, most notably International, a large truck manufacturer, have incorporated security into the HSE department, giving it a broad mandate for business continuity and protection of the enterprise. (See Chapter 22.) Very recently, this trend has included environmental sustainability in the portfolio of the HSE department. Such departments, which are becoming increasingly common, combine environmental health concerns of the type discussed in Chapter 12 with broader issues of sustainability outside health, such as energy conservation, recycling, reducing carbon emissions, and green technology.

All organizations benefit from an explicit occupational health and safety policy that specifies their responsibilities to their employees and their employees' responsibility for safe work practices. A model occupational and environmental health policy, one that could apply to any company large enough to have a corporate medical depart-

ment, will confirm the company's commitment to protect the health and safety of its employees and of persons living or working near company operations and will assign responsibility for the actions and decisions required to maintain this commitment.

### **Labor-Management Relations**

Each employer has its own pattern of labor-management relationships, unionized or not. Sometimes there is a mixture of both, as when a union represents some workers but others in another division or certain jobs are not organized. The responsibility for protecting worker health rests squarely with management. In the non-unionized setting, management must fulfill this obligation with no less diligence than is required when it has a union looking over its shoulder. In either case, effective policies and procedures to protect employee well-being should be in place. These should be based on a moral commitment that protecting employee well-being is a fundamental principle of doing business.

Whether a plant, company, or agency is unionized or not becomes important to the role of the OHS, but it is difficult to generalize how interaction with a union affects day-to-day operations. When employees are organized in a plant, the union may negotiate for improved working conditions as part of collective bargaining. In general, an OHS that is the direct product of collective bargaining is vulnerable to abrupt changes in policy, to financial cutbacks, and to being used as a pawn or distraction in contract negotiations. The OHS may be mistrusted by management as an unwarranted benefit for employees or may be seen as a beneficent gesture or concession, demonstrating the employer's good will and responsiveness; this is not necessarily desirable, as it can degenerate quickly into a patronizing and paternalistic attitude that turns into sour resentment at the slightest problem or conflict. It takes real effort to preserve the OHS as neutral ground in such circumstances and requires a commitment from both sides. This commitment can only come from a mutual understanding of the OHS and its actual role in the workplace,

regardless of what either party would like it to be for their own purposes. In unionized plants, the local often intervenes in individual claims as well as advocating changes in the workplace and acting as a conduit for complaints regarding working conditions. The union representative may assist the employee in filing a claim, arguing an appeal, or researching the background to a complex problem. Unless allowed by a clause in the collective bargaining agreement, the union representative is not entitled to personal medical information (other than fitness to work) any more than the representative of the employer, but the employee is free to release medical information to whomever they wish and will do so if the union is taking up the case. A collective bargaining agreement, on the other hand, is a contract binding on all parties until it expires. If the contract stipulates that specific information is to be or can be released, that constitutes valid authority to do so.

The negotiated collective agreement between an employer and its union (or unions) sets the tone of the relationship and lists many of the rules of behavior for the workforce in a unionized plant. While the collective agreement in no way removes the employer's right to run its own affairs, it binds management and the workforce to a contract. These rules are in force for the life of the collective agreement, which includes an arbitration mechanism for settlement of disputes. Issues concerning the well-being of workers, individually and as a group, form an important and often controversial part of collective agreements. Sometimes the terms of the agreement contain only broad standards, but others set down very specific policies and procedures.

All such collective agreements insist upon some form of worker participation in programs that deal with worker health and safety. A common stipulation (especially in Canada) is the joint health and safety committee. The powers of these committees vary greatly, ranging from total control of health and safety matters to serving merely as discussion groups. They provide a valuable opportunity for management and workers to come together to solve problems in the workplace, if the participants can keep focused on occupational health and safety.

## **Occupational Health and Safety**

Nonmedical functions, such as safety and occupational hygiene, may be organized as separate departments within an organization or incorporated into the management structure in an integrated occupational health unit. Sometimes specific nonmedical services are obtained by contractual agreement from an outside supplier and simply coordinated by an internal manager. Regardless of what form of organization is selected, it is essential that there be clear lines of communication and a close working relationship between the essential nonmedical functions and the medical department. Otherwise, efficient prevention and problem solving are impossible, and solutions to problems are often postponed or resisted because of personal or political agendas.

Among large employers in government and industry, a distinction is usually made between “occupational health” and “occupational safety.” “Health” in this context is presumed to refer to occupational diseases and the control of chemical hazards in the workplace. Although it is recognized that there is substantial middle ground, the two are usually considered separate areas of authority and are usually under the responsibility of separate management units. Table 20.4 compares the general approaches of “health” and “safety” as they are divided in many organizations.

Health is the newer area conceptually, reflecting increasing awareness and regulation of chemical hazards and hazards intrinsic to industrial processes. Health functions are usually conducted by occupational (industrial) hygienists. The management unit may be headed by a hygienist or an OEM physician, although sometimes this area is combined with loss and liability control functions under the direction of a manager with an appropriate administrative background. Health is usually perceived as highly technical, somewhat arcane, and primarily an issue of minimizing the risk of an adverse effect or citation under government regulations by controlling levels of exposure in the workplace.

Safety is a more traditional concern of management and has arisen out of efforts to prevent accidents, respond to emergencies,

**Table 20.4.** Comparison between Health and Safety Approaches

Safety	Health
Numerous specific problems, “low tech”	Numerous specific problems, “high tech”
Macro-management cost-effective Workplace audits	Micro-management cost-effective Workplace environmental monitoring
Periodic health evaluation not effective	Periodic health evaluation has role
Diagnosis obvious	Diagnosis often subtle
Traditionally addressed physical and mechanical problems	More recently has addressed ergonomic problems
<hr/>	
Middle Ground	
Repetitive strain/cumulative injury Chemical burns Physical hazards of toxic chemicals Disaster planning Rehabilitation from injury Drug abuse Identifying risk factors underlying injuries	

fight fires, and monitor the presence and handling of physically hazardous exposures in the workplace. These include flammable and explosive materials, high-voltage electricity, open flames, compressed gases, and mechanical hazards, such as unguarded equipment, motor vehicles, ladders, and scaffolding. Safety issues are usually dealt with by safety professionals trained in short courses after spending time as workers or foremen. The management unit is usually much larger than a health unit and is more closely tied to production units than health units. Safety personnel usually have greater visibility on the shop floor and are better known to rank-and-file workers. Of necessity, parallel safety units must exist in every plant the employer has.

Conflict between health and safety units, and between both and a separate medical unit, is common and typically reflects inadequate resources given to each or issues of authority over the large number of issues that fall into the “middle ground” between the two (or among the three). A more conceptually useful approach may be to consider health and safety as two ends of a spectrum of issues (Table 20.4). This model seems to work well, in general, but is difficult to put into place when the past model of organization has kept the functions separate and reporting lines have become rigid. A major concern in merging units is the allocation of limited resources to one or the other area: who will gain and who will lose.

## **FACILITIES AND OPERATIONS**

OHSs operate on the level of acute and primary care or on the level of referral and specialty consultation, but usually not both. The reason is that the former requires a high-volume, clinical service–driven model that depends on efficient delivery of services and speed and accuracy in completing paperwork, especially for workers’ compensation, and is generally reimbursed by billing systems based on service codes. The latter is a high-value-added, low-volume, cognitively driven model (meaning that the physician’s expertise and knowledge are more important than clinical skills such as suturing) that requires a heavy investment of time in each injured worker and is generally reimbursed by invoice at an hourly rate. The two models are not compatible in a single business model and have very different implications for staffing, equipment, and spin-off of ancillary services. This section primarily addresses the high-volume, acute-care business model.

### **Location**

Locations best suited for OHSs are areas of industrial growth. In particular, lower-technology industries, such as assembly line manufacturing or automotive repair, tend to produce more work-related

injuries and illnesses than capital-intensive high-technology or automated industries. Labor-intensive industries, particularly those employing large numbers of untrained or partly trained workers with a high turnover, such as fast food operations, tend to produce more acute injuries. Older, lower-technology manufacturing industries are more likely to result in occupational injuries, while higher-technology injuries often present exotic problems in toxicology and ergonomics. Virtually all industries, including service industries with mostly desk jobs, generate large numbers of back complaints. Office operations and financial or information service industries require a larger proportion of preventive and educational services and may create a greater demand for health promotion activities. Particular locations will tend to attract one type of industry. Large employers will often be surrounded by many smaller ones providing support services and goods. In many cities, the downtown area is changing to a service core and manufacturing-related industry has moved to the periphery. Older health facilities serving the downtown area must consider whether the mix of services they have provided is now appropriate for their local industrial base.

Some healthcare organizations attempt to provide both personal and occupational healthcare in the same facility and locate in areas where there is growth in residential population. Residential population means little for OHSs because industry is seldom in the same neighborhood. Clinics or other facilities in a residential community are often in a poor position to provide occupational services to employers in the industrial parks or districts where those same potential clients and patients work. One clinic in San Diego, for example, was established in a commercial area on the periphery of a large suburb in a metropolitan area in a location selected to be convenient both for young families in a residential area of that same suburb and to industry in an adjoining city. The building was accessible only by car, separated by a hill from a local housing development and by enormous parking lots from a shopping center. Mothers with children and the elderly found the location highly inconvenient, even though on a map it appeared to be close to a con-

centration of homes. The major employer in the area never sent injured workers to the clinic because several physicians were within a shorter drive, even though this employer was on the same street. Professional marketing services costing thousands of dollars were engaged to promote utilization of the clinic but had little effect. After three years, this satellite had to be closed because it had served neither industry nor residents adequately. The location, having been compromised in order to satisfy both groups, satisfied neither since it was poorly located to serve the residential area and situated even worse to serve the occupational market. During its brief period of operation, several small industrial medicine clinics established themselves much closer to the primary market for OHSs.

## **Facilities**

The siting and design of the clinical facility itself should be undertaken with care and preferably designed by reference to a successful model. It is impossible to describe a design that meets the needs of all organizations. Advice from consultants in both architecture and occupational health and visits to other facilities are always good ideas before one commits to construction and purchase of equipment. This section, therefore, will concentrate on general principles of space allocation, access, configuration, interior fixtures, and equipment.

Medications to be stocked are the minimum needed for the symptomatic relief of common occupational disorders or personal illnesses affecting employees on the job. Although occupational physicians cannot substitute for family physicians in providing comprehensive care, it is not unreasonable to treat common complaints in order to prevent unnecessary time off work. It is usually not cost-effective for occupational health clinics to maintain clinical laboratory apparatus beyond a bench centrifuge, microscope, and simple office testing supplies unless they are remote and must function as a self-contained infirmary. It is usually not cost-effective to acquire radiologic apparatus unless the industry is one at high risk for serious injuries and the site

is a large one. Maintaining quality assurance for such services can be a serious challenge when the volume of cases is low and there is a temptation to raise revenue, which increases costs to the customer (the employer or insurance carrier) by unwarranted utilization in an effort to justify use of the equipment. Usually, a nearby clinic or hospital can provide the support services needed and indicated at a lower unit cost, with a better guarantee of quality assurance and without an incentive to over-order tests.

An OHS should be able to provide a realistic cost-benefit analysis for the organization of which it is a part or that uses it for healthcare services.

## **Staffing**

The physician cannot be truly effective in managing an occupation-related case in isolation. Occupational health nurses are the essential occupational health professionals in most employer-based facilities and can deal independently with many cases. A complex case absolutely requires the participation of other professional experts. These may include occupational health nurses, occupational hygienists, audiologists, toxicologists, epidemiologists, engineers, radiation physicists, and many others.

As with all medical facilities, the staffing of an OHS depends on the mission of the organization and the health patterns in the industry and community. The number of patient encounters expected per year is a significant determinant of the size of the service overall and its projected capacity. The type of injuries or illnesses to be expected determines what special equipment should be readily available and what specialty coverage should be arranged. Close proximity to a hospital or large group practice, with the potential to receive referrals to specialists, makes a big difference in staffing; but if there is no such institutional support in the area, easy access on a timely basis to specialty services may require that the OHS have its own designated staff or working arrangements with local consultants. The administrative responsibilities of the service and budgetary constraints shape

the nonmedical staffing of an OHS. The anticipated peak utilization of the service and variations through the year dictate short-term hiring and whether special services (such as audiometry) are contracted out rather than performed in-house.

As with all medical facilities, the staffing of an OHS depends on the mission of the organization, the health risks in the industry, and access to supplemental assistance in the community. Some factors to consider are:

- Number of patient encounters expected per year
- Type of injuries or illnesses to be expected
- Degree of organizational autonomy enjoyed by the service
- Administrative responsibilities of the service
- Budgetary constraints of the organization
- Anticipated peak utilization of the service and variations through the year

Peak needs should never be the basis for full-time staffing projections, as they burden the operation with redundant staff during most of the year. Peak load can be reflected in the budget for employing part-time personnel for assistance during these times. Seasonal industries usually schedule periodic health surveillance during their lightest months, to minimize disruptions to production.

Staffing formulas for OHSs without particular physical or chemical risks have followed a rough rule of thumb for many years, despite economic changes. A plant physician on at least a part-time basis becomes cost-effective for about 1,000 employees and on a full-time basis for more than 2,000. An occupational health nurse is usually needed for more than 300–500 employees and for every additional 750–1,000 employees. A safety officer and access to medical care nearby is generally adequate for an employee population of less than 200. A full-time occupational hygienist will usually be required at about 500 employees in manufacturing, but part-time consultants may be used for much routine work and for special problems. These

crude guidelines do not take into consideration the increased risk and special hazards in some industries. An occupational health nurse may be quite sufficient for several hundred employees in an office building, but an occupational hygienist may be essential for a small company in a high-risk industry. There is no simple formula for staffing an OHS, but for medium-risk industries the above rule of thumb will not be far out of line.

## **Record Keeping**

An OHS must maintain at least two types of records: personal health records and exposure records.

Medical records contain confidential information and should never be accessible to unauthorized personnel, including management. These records:

- Document significant exposures sustained by the worker
- Match the worker's health and fitness for duty and job requirements (see Chapter 18)
- Document the worker's health on entry for fitness to work determinations,
- Summarize the results of periodic health surveillance (see Chapter 5)
- Document compliance with regulatory requirements

Exposure records are the results of environmental monitoring or group data on personal exposure; they are not necessarily confidential unless they identify individuals. The U.S. Department of Labor issued a ruling in 1988 on access to medical records that requires that exposure records be kept for all workers, even those exposed to little or no hazard in the usual work environment. Workers have access to their own exposure and medical records, with copies to be provided free of charge.

Both types of records are normally the property of the employer or whoever caused the record to be created in the first place. However, workers must consent to the release of any information that identified them as individuals, even to their own union. Anonymous data (with all identifiers deleted) and group exposure data may be shared but not personal medical records, which cannot be individually reviewed by any party other than the worker or an authorized health professional; management representatives are not entitled to view the record, even if identifiers are removed. Only the worker can authorize release of health information from the medical record and sharing of the information with other parties. An exception to this general rule is when records are subpoenaed for litigation or become part of a disputed workers' compensation claim or are requested under the legal authority of OSHA or NIOSH.

Records must be retained at least thirty years after termination of employment to permit review in case chronic or latent health problems are identified. These records should always be transferred to a responsible recipient or government agency if the employer or clinic goes out of business. If there is none, the records are to be transferred to the Director of NIOSH in the United States. This is probably rarely done.

Certain exemptions to the record access regulation have been made. Research is permitted, but individual identification of workers is not allowed without explicit permission. Employee assistance records are considered separate from the medical record if they are maintained in a different file; they remain confidential but are not subject to the record retention requirement and do not have to be disclosed when the medical records are subpoenaed or requested by OSHA. Also, employees who leave the employer after less than one year employment may be given their medical record to take with them, and copies do not then have to be retained and stored by the employer.

It should be noted that chest films are considered integral parts of the medical record, subject to the rule of thirty-year retention after termination. Old films therefore cannot be processed to reclaim the silver until the retention period is over.

Employers are responsible for ensuring that the procedures are followed and requiring healthcare providers outside their organization to follow these same regulations. In the present climate of increasing litigation, it is wise to be prudent in obtaining written authorization for all transfers of confidential information.

## **Marketing**

In the current competitive environment, marketing is necessary to preserve financial viability, but is also essential in order to respond in a timely manner to clients' needs. This second role of marketing is just as important as salesmanship. Marketing must be considered not only as a means of promoting a product or service but also as a means of determining what the consumers of that product or service require and adapting to the needs of the consumer. Applied to occupational medicine, marketing means listening carefully to understand the needs of the "consumer" of occupational health services, which is usually the employer, and providing services of high quality and convenience for benefit of the recipient of care, which is the worker.

Services should be designed to satisfy the needs of employers and the expectation of workers. Healthcare facilities and individual physicians providing occupational health services must accommodate the needs and preferences of injured workers in order to retain their share of the "market" of customers, the employers who send injured workers to them, and clients—the workers who choose them when seeking medical services. If employers perceive the waiting periods as excessive, and workers (clients or patients) come back dissatisfied, the facility will quickly lose customers (employers).

Occupational health services should not be used as a "loss leader" offered for the purpose of expanding the patient base for personal healthcare. The economic viability of OHSs should be the business reason for establishing a for-profit clinic or satellite. This economic viability can be guaranteed in part by establishing agreements, either by an informal letter of agreement or formal contract, with local employers to utilize the services of the new satellite on a trial basis.

Satellites that are not economically successful within a reasonable time period should be closed and the resources used elsewhere; attempts to repurpose them, such as converting family medicine offices to occupational health centers, usually do not work.

### ***Marketing Principles***

Employers entering into contracts with medical facilities for the provision of basic care select the provider on criteria quite different from what would be applied by a patient seeking a relationship with a personal physician. Six principles underlie the marketing of an OHS and distinguish it from marketing for general health services:

- The facility should be located where the workers work, not where their families live.
- The services provided by an OHS are sensitive to price; perceptions of quality differentiate providers within a price range.
- The facility should adapt to local needs and provide good service to local employers and their workers.
- A mix of large and small client employers is more stable than reliance on a single employer, however profitable.
- Occupational and personal health services should be kept strictly separate.
- The services provided by the OHS must be delivered in a timely fashion: waiting time must be minimal, turnaround time for reporting and paperwork must be quick, and reporting to the employer regarding fitness for duty should be timely.

The first rule is to locate where the workers work, not where their families live. To build up an OHS, medical care must be taken to the worker. The workers will not be sent long distances by the employer for care if alternatives are closer. Some studies suggest a limit of about three miles in urban areas. It is highly unusual for employers to send their workers with minor injuries across town just to be seen at a

clinic with a good reputation. As well, any facility that intends to provide U.S. Department of Transportation (DOT) commercial driver evaluations on a large scale had better be located on a main highway with high visibility, good access, and nearby parking for trucks.

Employers considering referral to an OHS are sensitive to price more than to quality. Medical services are exceedingly difficult for the lay person to compare in terms of quality. Employers are more likely to select a provider of medical care on the basis of cost and convenience of location and hours rather than attempting the difficult task of comparing medical capability and credentials of the staff. Employers, especially small enterprises, are persuaded that most routine services, such as acute care for injuries and pre-placement evaluation, do not require extensive specialty care and are within the capability of any qualified physician. Rightly or wrongly, adequate is perceived as good enough and a reputation for excellence counts for little among managers if the injured worker is not seen as requiring fancy care. This attitude also holds true for periodic health evaluations (except for executives), workers' compensation management (until a company has a difficult case), and other forms of fitness-to-work evaluation and applies to rehabilitation services. It does not apply to specialized care for occupational disease, independent medical evaluations, or consultation for preventive services. Within a price range, however, employers will generally choose a provider with a superior reputation or higher visibility to start with but will experiment with new or unfamiliar providers if they can be convinced that time lost will be minimized, case management is more focused, and that the management of workers' compensation paperwork is expedited and taken seriously.

The third general principle is that the OHS must adapt to local needs. As downtown industry is replaced by a clean, white-collar office and service workforce, the demand for specific occupational healthcare services will be far different. Expansion of an existing clinic in the downtown area may be less advisable than, for example, the development of a satellite in a rapidly growing suburb from the standpoint of caring for injuries. A downtown location may make

sense if prevention-oriented and health promotion programs are added to the range of services. The type of industry in the area will matter a great deal in generating demand for services, but some problems, such as back pain, are important in virtually every industry and office. Nothing creates dissatisfaction faster than services not being available as described. Services that are being marketed must be in place before they are heavily promoted or the facility will rapidly lose credibility among local employers.

The fourth general principle is to seek a mix of contracts or client employers. Although it is desirable to have big clients to ensure a high volume of services, it is dangerous to depend on a few big contracts that could disappear overnight, particularly those representing employers in the same industry subject to the same market trends. Successful OHSs usually develop a mix of large and small businesses in a variety of industries. Small business, of course, is “big” business in the aggregate and is just as important as big business itself in supporting a stable and profitable OHS.

The fifth principle is that the OHS must be separated from facilities providing personal healthcare services. The flow of occupational patients should not be mixed with family or primary patients. The expedited, time-sensitive flow of an OHS meshes poorly with primary care for the family. When occupational medicine is mixed with primary care, injured workers wait their turn behind patients awaiting general medical care, crying infants, and elderly and pediatric patients, who are often taken before them out of courtesy because most workers are or look healthy. This causes unavoidable delays and long waiting lines (which both workers and employers dislike), results in what employers perceive as lost productivity, and often promotes an unfortunate attitude on the part of the staff that the patient is “just workers’ comp” (in years past the phrasing would be “just an industrial”) and can wait. There is no equitable way to get around this problem in a shared facility. Giving priority to workers simply does not work in a clinic situation. Patients who are waiting to be seen, especially with children, feel slighted when workers with minor injuries or a seemingly administrative reason for the visit, such

as a pre-placement evaluation, bypass the line. Except for small, individual, private practices, therefore, it is best to insist on a strict separation whenever possible between OHSSs and general medical services, preferably by putting the two in separate facilities altogether. There are really no effective halfway measures. At a minimum, waiting rooms should never be shared.

Occupational medicine practice is fundamentally about timely care within a system that requires timely reporting. Employers need and expect fast service, rapid reporting, and adequate quality of care, which, as they see it, is reflected in accurate and complete paperwork that is filed without delay. Time off the job because of injuries or routine evaluations costs the employer money that the smaller business in particular can ill afford to lose. Uncertainty over how long the employee will be away from work compounds the loss by causing confusion, inefficiency, and difficulty meeting schedules.

### ***Representation and Communication***

Direct face-to-face contact between marketing representatives and the employer's representatives is absolutely required for effective marketing. Most physicians are not good at this, and their time is much too expensive to take on this function, beyond initial visits to a new client. Physicians are often too threatening and imposing to nonphysicians to be effective marketing agents. They will usually be treated with less candor compared with a nonphysician. A good marketing representative, on the other hand, can visit at regular intervals to trouble shoot problems and will get a more candid view of problems or complaints. Properly trained and instructed, a marketing representative can serve as the eyes and ears as well as the human face of the healthcare facility, but caution must be exercised to keep the overenthusiastic types in line. It can be very dangerous—and expensive—to turn marketing responsibilities over to a contractor or independent representative who stays away for long periods without supervision.

Marketing representatives must be careful not to oversell the capabilities of the clinic and must not be too quick to agree to employer's requests for specific services since some of them may be ill-advised. For example, one marketing representative committed a clinic to perform routine back x-rays for screening purposes on new employees, a procedure that is not acceptable practice. The representative had spent considerable effort selling local industry on the need for this unnecessary service, and the clinic was put in an extremely embarrassing position. Experience has shown that the best marketing representatives for occupational health facilities are relatively conservative in appearance, appear professional in manner, have some health-related experience themselves, and are willing to work closely with the professional staff of the facility. Former pharmaceutical representatives are often excellent candidates.

Clear lines of communication with both employers and also with workers using the facilities ensure that needs are met in a satisfactory manner. Web sites are, of course, essential for marketing and communication and absolutely necessary if an employer seeking a service is to find the facility. However, they are not enough. Web sites are passive media, accessed only if the employer is looking for them. Other media are required to bring the facility to the attention of local employers and to establish its niche in the community.

One way to promote such communication is by circulating to the persons responsible for workers' compensation and occupational health matters in each employer's organization a newsletter giving tips on health that can be reproduced for employees, news of the healthcare facility and its personnel, and items of significance to local industry in occupational health. This gives the healthcare facility visibility and promotes a more personalized relationship with the employer that makes it more difficult for the employer to break abruptly without explanation.

An OHS does not imply the same commitment to continuing care as a family practice or a health facility serving personal healthcare needs. The management of occupational injuries and illnesses is usually episodic, and only some cases require continuity of care.

Hence, there is a less personal commitment to healthcare than one expects in a primary care setting, such as family practice. If employers are not satisfied, they will often simply send injured workers elsewhere. Employers can and do shift their preferences for providers and feel no obligation to explain why. Following up with employers and obtaining feedback, while respecting the limits on their time and patience, are an important part of the art of marketing.

### ***Marketing Strategy***

A marketing effort dedicated to OHSs should have its own budget, a strategy to reach a broad range of local employers, and a sincere effort to provide empathic and satisfactory service to workers. It should be responsive to its clients in the form of facilities, staff skills and training, and quality assurance that is consistent with its claims. This cannot be achieved when the waiting room is shabby, parking is inaccessible and expensive, clinic hours are limited, and the staff does not behave in a professional manner. The facility has to “walk the talk” if marketing is to yield sustained results.

A sound marketing strategy requires knowledge of the range and levels of services that can be provided matched to the potential customers for these services, the employers. (The employees are the clients, if they are well, or the patients, if they are injured, but are not customers unless they choose the OHS themselves.) Table 20.5 lists various services that might be provided by a large OHS. The marketing strategy should be based on a realistic assessment of economic activity in the area, one that takes into account all industrial sectors and not just the most visible or prestigious ones, and it should utilize current profiles of employers across a wide range of sizes and sectors, not just the largest or most prominent employers.

Figure 20.3 is a diagram known as a “marketing cube” that allows a three-dimensional representation of the possible combinations of levels of service, range of services, and users or “consumers” of these services. A marketing cube presents three dimensions of OHSs: range of service, users of service, and level of service. Each subdivision of

**Table 20.5.** Core Functions of an OHS

- 
1. Acute care for injured employees.
    - 1.1. Providing care on site
    - 1.2. Monitoring care given off site
  2. Pre-placement evaluations
    - 2.1. Assessing functional capacity to do the job
    - 2.2. Assessing need for accommodation under the Americans with Disabilities Act
  3. Functional evaluation of employees after hire
    - 3.1. Fitness-for-duty evaluations that assess the recovery and functional capacity of injured employees to return to work and what accommodations may be needed
    - 3.2. Impairment evaluation for injured workers with permanent impairment and workers' compensation claims
    - 3.3. Certification of time off work for workers with a nonoccupational illness or injury. (This is often performed by other physicians.)
  4. Review of workers' compensation claims for causation.
  5. Periodic health surveillance of employees exposed to a particular hazard such as noise, chemicals, dusts, or radiation (typically takes the form of a medical examination, often conducted annually)
  6. Investigation of exceptional hazards, disease outbreaks, unusual injuries, fatalities, or other emerging issues
  7. Prevention, health promotion and educational programs designed to enhance the health of employees and to increase productivity
  8. Management of the health problems of employees on-site, to reduce absence and disability
  9. Advice and consultation to management on issues of health, health and workers' compensation insurance, and regulatory issues in occupational health
  10. Disaster planning and emergency management on or off site
  11. External communications on health issues, as with local public health agencies and local physicians
  12. Managing relations between the organization and local hospitals and the medical community
  13. Employee assistance programs, for employees with problems involving alcohol and drug abuse or other addictive behaviors, such as gambling, that interfere with work

*(Continued)*

**Table 20.5.** (*Continued*)

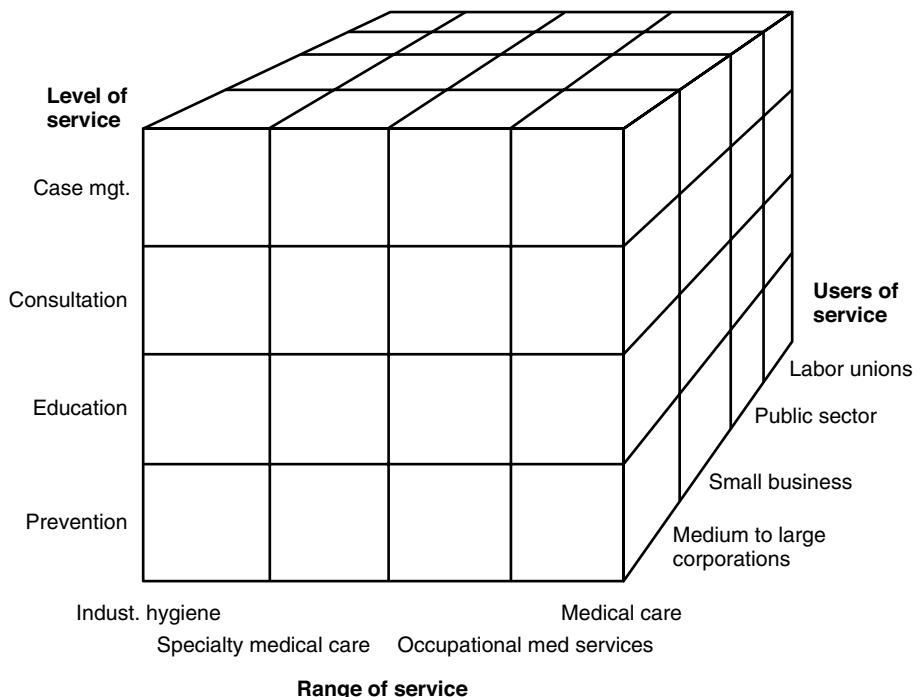
14. Executive wellness programs, such as special medical evaluations or monitoring health problems among senior executives

Larger and more complex organizations may also involve the occupational physician in managing environmental risks, product safety, contracting for health services, representing the organization in industry-wide health activities and proactive programs for preparedness, risk management and other senior management functions.

the cube represents a particular level of a particular service provided to a given category of user. Presenting these three dimensions in one illustration helps one to visualize the possibilities and to identify opportunities for growth.

Each level of service can be matched with each service type in Figure 20.3. The medical services may deal, for example, with treatment of individual cases or preventive services to groups of workers. These are the most common types of medical services, but opportunities to provide consultative and educational services are often overlooked. While physicians' expertise may not be well utilized on a cost-effective basis by providing only educational programs, highly professional programs designed with physicians' input can be provided at reasonable cost by health educators or nurses and are very popular among some types of employers and groups of workers. Industrial hygiene services are usually provided on the case management or consultation level, but a market can be created for periodic assessments for purposes of prevention and to ensure compliance with government regulations.

The users of the services, shown along the base of the cube in the front-back dimension in Figure 20.3, may include larger businesses, small business, public agencies, and, potentially, labor unions. Individual workers do not constitute a market for OHSS in the same way that they and their families are a market for personal healthcare. The "consumers" of healthcare are those who use the system and make the choices. In the occupational healthcare system, it is usually the



**Figure 20.3.** Marketing “cube” for OHSs.

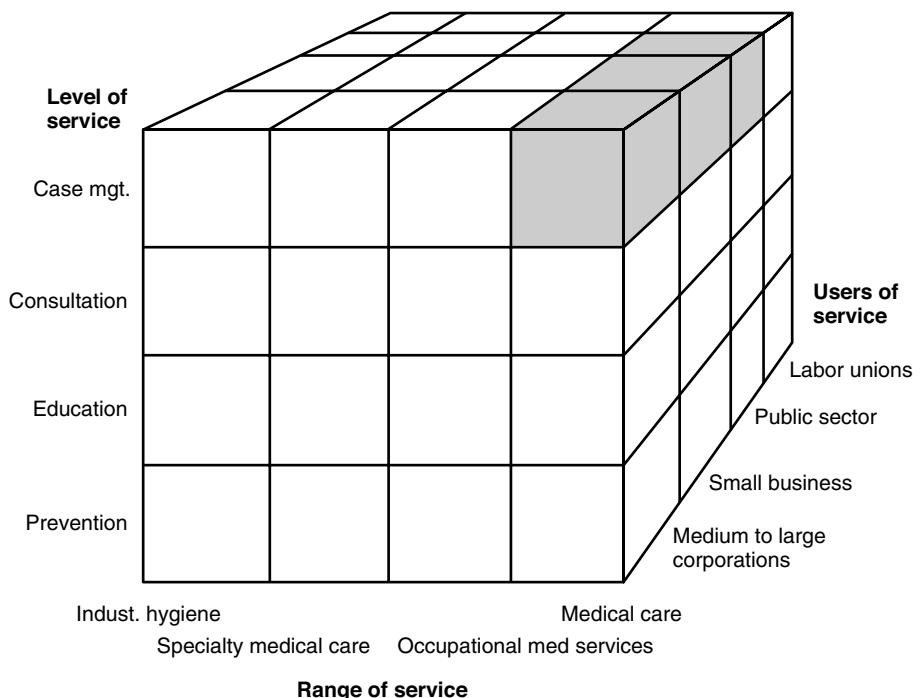
employer who makes the initial choice and who purchases (directly or through workers’ compensation) health services on behalf of the worker. Even when an individual worker changes physicians or seeks care for an occupational health problem from his or her own doctor, the system constrains the choice by allowing only a limited number of charges and refusing to pay for unauthorized referrals. In marketing OHSs, therefore, the essential target is usually the employer. As a practical matter, the workers’ needs must always be met, but the employer’s needs must also be reasonably satisfied or the relationship between provider and client may be brief.

The level of services, on the vertical axis, represents a continuum from direct case management to prevention. Diagnosis, treatment, rehabilitation, and follow-up are familiar as the medical model, but the approach of direct intervention also applies to the management

of specific problems that arise in the workplace, such as hazards that have been identified or clusters of health problems suggesting that a hazard must be searched for. Consultation is problem evaluation and solving that requires less direct intervention but particular insight and expertise, not only into the problem but also into the needs, motivations, and resources of those asking for the assistance. Education includes not only formal training sessions but opportunities to increase the awareness and sophistication of clients so that the services used are more highly valued. Most fundamentally, prevention is the foundation of sound occupational health practice.

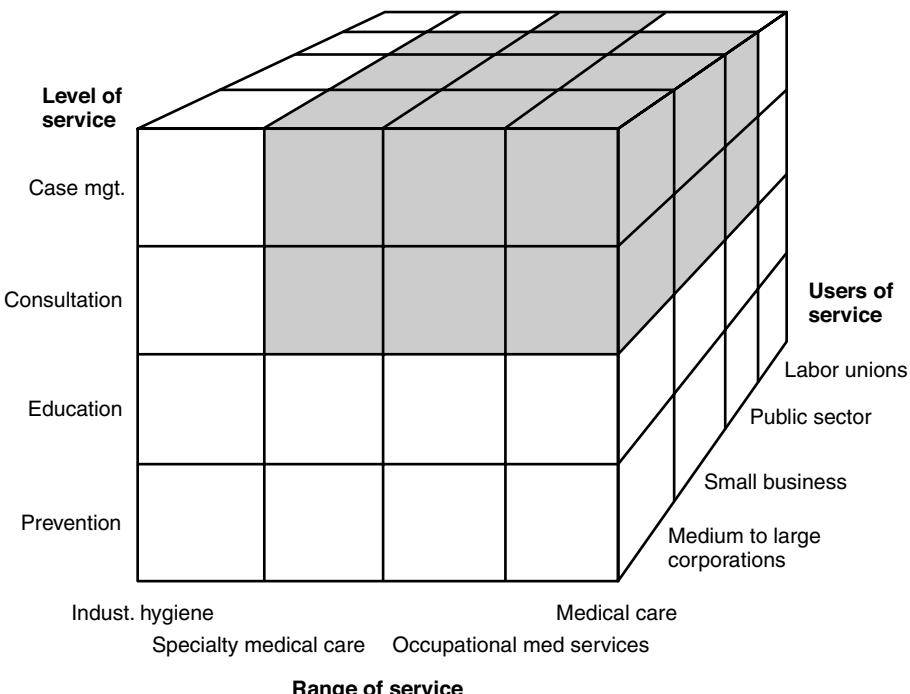
The range of services, presented in the horizontal dimension, is the easiest to conceptualize. “Medical care” includes routine health services, including acute care, periodic health surveillance, and possibly services performed on-site in the employers’ facilities. “Occupational medicine services” in this context refers to the specific services such as periodic health surveillance and workers’ compensation management that require knowledge and understanding of the occupational healthcare system. “Specialty medical services” refers to an advanced level of care in medical diagnosis and treatment, such as dermatology or toxicology. “Industrial (or occupational) hygiene” is the recognition, evaluation, and control of hazards in the workplace and is performed by industrial (or occupational) hygienists, a specialized profession, as described elsewhere in this chapter.

The most basic level of care is within the capability of most primary care physicians and clinics and is marketable to local area employers. A clinical practice will compete on this level primarily on price (because employers assume that basic medical care can be provided by any qualified medical practitioner), convenience, short waiting times, and prompt completion of paperwork. Figure 20.4 illustrates the marketing cube for this primary level of care. From a business point of view, this is a high-throughput, relatively low-margin operation, and its growth potential depends on local employers and the profile of industry.



**Figure 20.4.** Marketing cube for basic clinical services, such as might be provided by a physician's office or clinic without an OEM physician.

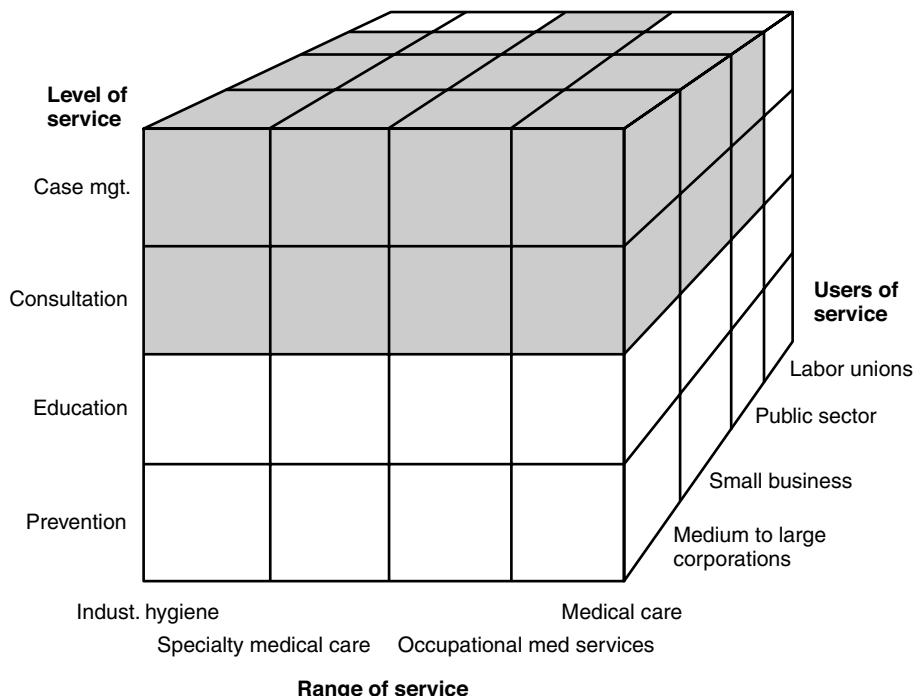
“Occupational medicine services” include specialty care provided by occupational physicians, usually emphasizing occupational disease management, toxicology, independent medical evaluations (IMEs), consultation by referral, and medicolegal services. Specialty medical care is rendered on a referral basis for special problems. These higher-level services are marketable to a wider range of clients and sometimes, especially in the case of medicolegal services, to insurance carriers, workers, and employers out of the immediate area. These are services that, in a business sense, add value and lead to a practice that may have lower through-put but equal or greater revenue because of higher fees commensurate with the higher level of service. To sustain such an operation, however, requires qualifications and credentials, such as board-certification in occupational medicine, that indicate



**Figure 20.5.** Marketing cube for primary care and specialty services in OEM.

competence and, especially for IME and medicolegal work, that establish medical credibility when there is a dispute, which is always. Figure 20.5 illustrates the marketing cube for a well-established occupational medicine specialty practice.

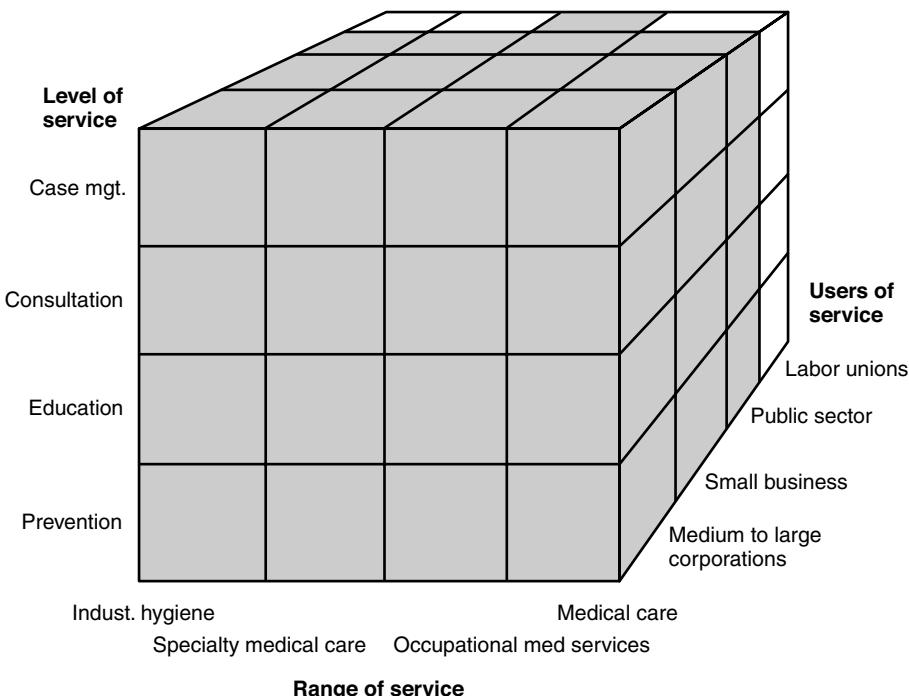
“Integrated OHSs” combine medical practice with occupational hygiene or other support services in order to evaluate the workplace and to assist in problem solving. They are necessarily team operations. Occupational (or industrial) hygiene services must be provided on the employer’s site (usually with a laboratory as a base of operations elsewhere) but are not provided by physicians. There are fewer examples of this type of practice than others, but they are especially appealing to large employers. Occupational (industrial) hygiene services are usually provided by consultants hired for the purpose by



**Figure 20.6.** Marketing cube for integrated OHSs, including occupational hygiene.

the employer if they are not available in-house. This split between medical and engineering services may not be logical from the standpoint of resolving the problem, but it reflects the different professional roles of the physician and the engineer. Where industrial hygiene services have been offered by clinics, they have often been undervalued or subordinated to the medical services despite their critical role in evaluating and controlling hazards.

Because they require more people and specialized equipment, integrated OHSs carry a higher overhead than other forms of occupational medicine practice. An advantage is that such practices can expand the scope of their operations without adding more physicians. Figure 20.6 illustrates the marketing cube for an integrated occupational medicine and hygiene.



**Figure 20.7.** Marketing cube for comprehensive OHSs.

“Comprehensive OHSs,” as illustrated in the marketing cube in Figure 20.7, are full-service practices, usually consulting operations with close ties to clinics or academic groups (such as member clinics of the Association of Occupational and Environmental Clinics), that can provide employers with prevention, education, and health promotion programs as well as hygiene and clinical services. The trend in such operations in the private sector is to establish national presence and visibility and in all sectors to work through networks rather than to house all functions in one office.

Occupational health clinics providing services for multiple employers usually emphasize basic medical care and occupational medicine services. Large group practices are also in a strong position to provide specialty medical care.

Outreach to employers is not necessarily reassuring to injured workers, who may fear that the relationship between employer and healthcare provider is too close. There should be a concomitant effort to demonstrate to the injured worker that the service is accessible, caring, worker-centered, and respectful of privacy. This is only right, but it is also good business. If they have any alternative, employers are unlikely to continue to utilize an OHS if they get complaints from their workers or hear stories about inefficiency and wasted time. Thus, although the primary marketing target has to be employers, every effort should be made to ensure that workers are satisfied with their care and feel well treated.

Labor unions sometimes have their own insurance plans and sometimes have their own medical consultants or services. They are also potentially sources of referrals, often for disputed or difficult cases. Members tend to rely on unions for advice on medical referrals for occupational problems more than on their personal physicians. Unions sometimes contract for educational and preventive services for their members but usually prefer to work with academic or nonprofit organizations.

Marketing is a particular problem among OHSs sponsored by hospitals and large group practices because the marketing effort is almost always subordinated to an overall institutional marketing plan. Hospitals and group practices obviously want to maximize the yield of their overall marketing budget. They are usually not enthusiastic about a separate marketing campaign for their occupational health referral clinics. However, advertising in the community for personal healthcare services is not the same as persuading workers' compensation carriers and employers that an OHS is the best place to send difficult cases, will expedite delivery of care, and will reduce waiting time to a minimum. This cannot be achieved if the OHS is combined with primary care or other specialties (especially pediatrics) and shares a waiting room with general patients (and especially children with contagious upper respiratory tract infections). Most administrators, however, expect the OHS to start small and cheap and upgrade only as it demonstrates its viability from a business perspective.

### **The OHS in Hospitals and Large Groups**

OHSs that are sponsored by hospitals, large multispecialty medical groups, managed care organizations, and integrated healthcare systems have common problems.

Most hospitals and group practices do not allow acute or primary care OHSs to develop properly. There are several reasons for this. Most hospital administrators have similar attitudes:

- They expect a high-volume, high-revenue-generating clinical service but will not design facilities nor staff, automate, or streamline procedures to support efficient operation. (This failing is most common in hospitals.)
- They expect the occupational medicine service to conform to a familiar model of service delivery, such as a high-value consultation practice (e.g., cardiology). (This failing is most common in group practices.)
- They do not recognize or acknowledge occupational medicine credentials, the nature of the specialized occupational medicine services, or the need for support services involving hazard assessment and control.
- They want to protect their existing emergency room or urgent care facilities from competition.
- They see paperwork flow as a secondary issue rather than a primary, essential, deal-making or deal-breaking part of occupational healthcare service.
- They are reluctant to set up a special administrative system for case management for OEM practice.
- They resist purchasing software because they want their in-house system to be used. (General healthcare software is poorly suited to occupational healthcare services, which specialized packages to be efficient and productive. There are several good products available.)

- They choose not to make an investment because they have been led to believe that occupational medicine is a cash cow that can fund itself from revenues.
- They price healthcare services so that they won't undercut their fees for general healthcare, thus pricing these services out of the market for employers.
- They do not market the OHS, or they do so only in combination with their general marketing, which is not aimed at employers and is mostly irrelevant to them.

Hospitals and large group practices are usually reluctant to hire staff on the basis of experience in workers' compensation, or those who are qualified for special occupational health services, such as "medical review officers" (MROs) or provision of commercial driver examination services (following United States Department of Transportation requirements; see Chapter 18), or staff with the specialized skills to manage workers' compensation cases well. Instead, they usually assign untrained clerical staff or nurses and resist efforts for them to be trained to a high level of expertise. This is not just to contain costs. It is also so that the staff can be reassigned again in the future without resistance and because using higher levels of expertise would lead to expectations for increased pay. Hospitals and group practices also expect that trauma care will generate referrals to surgery as much as possible—which most occupational injuries, involving soft tissue, do not need—and will use in-house rehabilitation, which may not be convenient for the injured worker.

The essence of responsive occupational medicine practice is case management and navigating the patient through the process (setting goals, coordinating care, tracking, monitoring progress, assessing fitness for duty). Hospitals are very reluctant to invest in the systems, software, and personnel to do this because there is no counterpart in other parts of their system and because it would require investment in up-front costs, before revenue is recovered. Hospitals and group practices almost invariably expect the OHS to fund its growth from the balance of revenues after overhead and almost never invest in it.

Hospitals and groups are reluctant to grant discounts in pricing clinical services such as audiometric examinations and chest films, in part because they do not want to set a precedent of departing from the fee schedule for other clinical services and in part because they want to maximize their overall revenue. However, a diagnostic test is not a screening test, and the cost is usually unacceptable to employers for pre-placement evaluations, periodic health evaluations, and screening. The result is that employers will usually go elsewhere and not return, if they have an alternative. Similarly, overhead does not reflect the realities of occupational health practice. Overhead calculations are made on the basis of expenses incurred by the entire hospital medical staff or the whole group practice, and the multiplier that is calculated from it (for example, a clinician may be expected to bring in three times as much in billings as his or her salary or "draw" for the year) takes into account many expenses for equipment and support services that the OHS almost never uses. The result is an overhead burden that often makes the OHS noncompetitive with other providers for anything other than acute injury care.

Healthcare administrators often try to force the referral clinic into an inappropriate model of specialist care, such as cardiology. This is a particular problem for practice plans at teaching institutions, hospital systems with outpatient facilities, and multispecialty group practices, where managers think they know what a specialty clinic looks like and expect the OHS to fit into that model, which does not serve occupational medicine well. The typical case referred to a consultant specialist for clinical evaluation in occupational medicine is extremely time-consuming, and billing fees for clinical services often cover only a fraction of the overhead and personnel cost. The best an OHS emulating this model can usually achieve is to be self-sustaining on the basis of revenue generated from a mixture of cases that includes a large fraction of patients requiring low-intensity screening and preventive services that cost little to provide. However, unless the consultant is well known, high-value cases usually will not come to the OHS in the first place.

Although referral clinics and academic occupational medicine specialty clinics have these inherent problems, they confer advantages

to the institutions and providers that can far outweigh their economic disadvantages if they are used strategically. The clinic may be the cornerstone of a practice that brings in much more revenue from “high-value” cases, such as independent medical evaluations and medicolegal cases. However, these cannot be supported unless there is an outlet for clinical practice. Unfortunately, the accounting and tracking systems of hospitals and group practices, being service code driven rather than invoice driven, do not recognize this income stream and so it is invisible to hospital and group administrators.

Hospitals and group practices expect that specialty clinics will spin off large billings in procedures, clinical tests, and consultations. Occupational health referral clinics rarely can do this because patients have often been seen by many physicians and have been extensively worked up before they are referred to clinic and by then the primary challenge is not treatment but causation analysis, impairment, and case management. OEM physicians do not perform expensive procedures, but they do bill at relatively high rates per hour for specialized services, which the hospital or clinic, again, cannot recognize through a conventional billing system. These fiscal constraints can be overcome by developing a mix of cases, basing such services at institutions such as teaching hospitals where physicians are on salary and overhead is subsidized, and associating a specialty clinic with a medical facility that also deals with occupational injuries at the primary care level.

Hospitals and large groups are logical homes for OHSs, but the services may not do well unless the hospital or group allows them to develop and grow according their own logic and in response to occupational medicine, not by conforming to a generalized template for healthcare.

## **SERVICES TO SMALL ENTERPRISES**

Small enterprises are employers that operate with a small number of employees relative to their industry sector. Small business represents by far the largest number of companies in the private sector and as a sector employs the most Americans and Canadians. The U.S.

Small Business Administration reported in 2008 that by its definition of small business, such enterprises constituted 98.2 percent of American businesses, employing over half the nonfarm workforce, paying 45 percent of the total payroll, generating up to 80 percent of new jobs and 28 percent of export value, and employing 40 percent of engineers, scientists, and other highly technical personnel. This is true not only for the United States and Canada but in every economy around the world, regardless of economic system. Of course, these statistics, while emphasizing the scale of the enterprise, mask the great diversity of small enterprises, from high-technology start-ups, to franchises of fast food retailers, to the corner dry cleaners. The defining issue in common among these enterprises is that, whatever other business advantage they may have, they are too small to achieve economies of scale in support services that are peripheral to their main product or service line, such as occupational health and safety.

For small companies, of fewer than 50 employees, the cost of regulatory compliance (occupational, environmental, and socioeconomic) is estimated to be seven to ten times as great per employee as for larger firms of 50 to 250 employees. A similar but lesser differential probably applies to companies with fewer than 500 employees compared to companies with more than 500. Industries with fewer than 100 employees rarely employ full-time occupational health professionals and depend instead on either part-time healthcare providers (physicians or occupational health nurses) or on community-based services. Many of the smallest companies are owned and operated by families or are vehicles for a single individual. Companies with 20–250 employees have a higher injury rate than either smaller or larger corporations. Thus small enterprises have a particularly acute need for occupational medicine services delivered at reasonable cost.

Large corporations have several advantages over most small firms in complying with good occupational safety and health practices, among them a much larger pool of technical talent to correct hazards, established in-house medical programs, favorable cash-flow and capital margins allowing resources to be put into occupational health and safety, and the capacity to assign responsibility for full-time manage-

ment of safety and health programs to specific employees. The small firm is almost invariably working on a much narrower margin and by definition has far fewer resources in capital, personnel, and facilities at its command. Because of its obvious disadvantages, the small enterprise requires special attention if it is to attain improvement in occupational injury and illness rates.

The direct benefits of prevention-oriented OHSs for small enterprises are also proportionately greater than for large concerns. Few small enterprises can absorb the consequences of absence or disability of a trained employee or manager without the loss being reflected in performance. In very small enterprises the workforce may be made up mostly or entirely of family members. One might think that small enterprises would demand preventive services up to the limit they could afford. In practice, this is rarely the case.

The decentralized nature of small enterprises makes it difficult to serve their needs in a coordinated fashion. Practical limits on cost and local accessibility make serving small enterprises logically difficult at a consistent standard.

In the United States, services to small enterprises are incompletely covered, both geographically and in terms of service mix, but are more readily accessible than elsewhere because of widespread outsourcing of occupational services in the 1980s. This seeded communities with providers and facilities that were previously available only in-house at larger companies. Most of these services, including both medical and hygiene services, supported themselves primarily but not entirely with contracts to larger employers, leaving enough capacity and creating a financial incentive to provide services to small enterprises.

At various times, governments have supported pilot projects to provide small enterprises with centralized OHSs. The best known of these demonstration projects is the Slough Industrial Health Service, established in 1948 in the United Kingdom to serve the smaller enterprises clustered in an industrial park called Slough Estates, in Berkshire. This prototype was highly successful in its day but required an ongoing subsidy to maintain a full range of services, including occupational hygiene. It continues today under the name Corporate

Health. Sweden developed a network of occupational health centers covering the entire country as a result of a labor-management agreement. In the early 1990s, the entire system was a political casualty of a more conservative government. In many countries, such as Kuwait, occupational health clinics are supported by the government. Government-sponsored occupational health centers have not been developed in the United States or Canada except in Quebec, where they are incorporated in the community health center network and OHCOW in Ontario, which is more often used for consultation than primary services.

The logistical problems of providing consistent service to small enterprises are formidable, but the potential rewards more than justify the effort. Most of society's economic activity and employment is to be found in small enterprises. Developing services for small-scale industry is therefore an important strategy to improve occupational health standards in general, and the effect of even small improvements consistently applied will be multiplied by the larger numbers of workers involved. Small industry must now cope as best it can and usually must turn to resources in the community to provide assistance in solving its problems.

Small enterprises have many competing priorities and often perceive occupational health and safety as a regulatory burden imposed on them rather than as an integral part of operating a business. Many managers in small enterprises have no clear idea of the hazards of their workplace or think only of one or a few highly visible hazards. This should not be surprising, because their interest and expertise is in business, not health. Invisible threats, such as cancer-causing chemicals, tend to be lost among urgent day-to-day matters of business. A handful of these managers, especially owner-managers, take the attitude that acceptance of some risk is a necessary part of success in their business and that if they have taken risks themselves, their employees should be prepared to do so as well. The great majority, however, are simply not knowledgeable about the cost-effective prevention of injury and illness in the workplace and would have little time to learn if they were aware. Provision of sound occupational medicine services to small enterprises requires the easy availability of

effective consultation to solve problems and education regarding the value and appropriate utilization of such services by small enterprise managers. It is not reasonable to expect all managers of small enterprises to solve their own problems in their own way. They must have help based nearby in their own community.

Managers and owners of small enterprises tend to be conservative in their use of OHSs, limiting use to the care of injured or ill employees or screening activities required by law. There is usually more ready acceptance of treatment and rehabilitative services than preventive-oriented services. Most perceive preventive or health promotion services with some skepticism, but the depth of this attitude depends in large part on the nature of the industry and the attitudes in the community. Occupational (industrial) hygiene, the evaluation of hazards in the workplace and the design of controls, is not a familiar concept to many small enterprises managers, who assume that they will incur an excessive cost. It should be a function of all services available to small enterprises to educate and guide managers in achieving greater cost-effectiveness through selective utilization of medical and hygiene services.

To the owner or plant manager, occupational health coverage is a consumptive rather than an investment cost, as it might be seen from the overall perspective of society. A heavy involvement in prevention or hazard control or health promotion may or may not forestall a future event from occurring; the cost is tangible, but the benefit is not. Furthermore, employees come and go, and a heavy investment one year may be lost with personnel turnover the next, particularly since the benefits are usually apparent only in the long term. In larger companies costs associated with disability and absenteeism may be reduced by company medical programs, but to a small company the costs of maintaining health services is usually out of proportion to the visible benefits. Although intensive safety control applied to companies with high injury rates does appear to reduce downtime, lower insurance premiums and workers' compensation levies, and improve morale, it is doubtful that managers of small enterprises that perceive them to be low in risk to begin with can be motivated to institute expensive in-house programs except when these are used as

benefits to attract exceptionally talented employees. Costs must therefore be kept as low as practicable when marketing to small enterprises.

There exists a severe shortage of the skilled professionals in occupational health and safety who are available to supply services, especially to small enterprises. Four specific measures have been proposed to overcome this problem:

- Increased training in occupational safety and health in medical school and expanded postgraduate continuing medical education in occupational medicine in order to increase the expertise of practicing primary care physicians
- Provision of clinic-based occupational safety and health services serving many employers
- Expanded use of nonphysician health professionals, particularly nurse practitioners
- Simplification of procedures such as standardized medical examinations for fitness-to-work evaluations appropriate to particular occupations or workers exposed to specific hazards to be performed by a physician untrained in occupational medicine

Outpatient clinical services for small enterprises have been provided for decades by private practitioners and more recently by group practices. These services have been largely inconspicuous, however, since they constitute only a small part of the activities of such practices.

## **MANAGED CARE IN OCCUPATIONAL MEDICINE**

Increasingly, primary medical care in the United States is provided by managed care organizations (MCOs), professionally managed groups of physicians and other healthcare professionals organized into corporate entities to achieve coordination and economies of scale. Physicians are subject to controls and restrictions on their practices

imposed by institutional procedures, clinical guidelines, limited pharmacopoeia, prior authorization, and utilization review. Increasingly as well, MCOs market their services to groups of patients defined by residence, employer, or other common characteristics.

Today, even services delivered outside of an MCO are likely to be reimbursed below billing on a fee schedule based on median rates or by a capitation allowance, with the financial risk accepted by the provider rather than the third-party payer. Because of intensifying competition and the increasingly risky financial situation in which medical practitioners and health providers such as clinics and hospitals find themselves, occupational medicine services are often used as a way to attract large groups of employees and their dependents as a patient base to enhance the financial stability of a captive health maintenance organization or preferred provider association.

Occupational medicine services, however, are not as driven by managed care as general medical care. Reimbursement rates for workers' compensation, for example, are now quite competitive with fee schedules for general medical services. The OEM physician also has the opportunity to engage in other, high-value services that have no equivalent in general medicine.

Fee-for-service, in theory, favors overutilization of services because the more services provided, the more fees the provider collects; in most cases, it is the physician who recommends a treatment, not the patient. Prepaid practice, in theory, favors underutilization because the less the provider must expend in costs to provide services, the more the provider saves from the prepaid amount. That is why third-party payers have increasingly introduced utilization review procedures to ensure that the care given is medically indicated and why health maintenance organizations, which are prepaid, conduct quality assurance reviews to ensure that patients are not going without needed care. In workers' compensation, the physician is compensated on a fee-for-service basis according to a fee schedule fixed by the workers' compensation board. The interests of the employer, the carrier, and usually the patient favor a speedy return to work. The idea is to further the interests of all parties by returning the patient to work promptly,

thereby reducing the losses to the employer from absence, returning the patient to his previous earning potential, and settling outstanding claims on the part of the carrier.

A basic difference between general and occupational healthcare is appropriate utilization. Frequent return visits and heavy use of rehabilitative services such as physical therapy, to a degree that would seem to constitute overutilization in general healthcare, may be quite appropriate in occupational medicine practice as long as they contribute to returning the patient to work as soon as the patient is able. The intensity of care is intended to further the interests of all parties: to return the patient to work promptly, to reduce losses to the employer from absence, to promote prompt settlement of claims, and to reduce long-term expense on the part of the carrier.

### **Liability Position of OEM Physicians**

Most coverage for OEM physicians is provided by the employer. For physicians in private practice, individual policies are available through the American College of Occupational and Environmental Medicine, which provides coverage for both liability for clinical services and errors and omissions for nonclinical services, such as consulting and medicolegal work.

Insurance companies in the United States and in Canada (specifically the Canadian Medical Protective Association) categorize physicians into risk categories, with as few as three and as many as nine classes. OEM physicians tend to be treated very differently by different companies. OEM physicians, almost uniquely, fall into either the highest or the lowest risk categories depending on their particular situation and their degree of formal training. The factors that are most often cited in determining whether an OEM physician would be considered low-risk are board certification in occupational medicine, whether one's activities are primarily of an administrative nature, the emphasis on preventive medicine in the practice, and whether the practice is included within a large organization. Factors cited as determining a high-risk classification include the lack of specialty

credentials, a preponderance of patient care responsibilities, emphasis on surgery or on trauma cases, and solo practice.

The most suitable type of coverage for an OEM physician depends, in large part, on whether one's practice involves clinical care in high-risk situations and to some degree on whether one is in solo practice (and thus relatively unprotected) or in an institutional setting.

## RESOURCES

- Guidotti TL, Cowell JWF, Jamieson GG. *OHSs: A Practical Approach*. Chicago: American Medical Association; 1989. [Reprinted as a "classic" in occupational medicine by Blackburn Press, <http://www.blackburnpress.com/>]
- Moser R Jr. *Effective Management of Health and Safety Programs: A Practical Guide*. 3rd ed. Beverley Farms, MA: OEM Press; 2008.

## NOTEWORTHY READINGS

- Cullen MR. Occupational medicine: a new focus for general internal medicine. *Arch Intern Med*. 1985;145:511–515.
- De Hart RL. *Guidelines for Establishing an Occupational Medical Program*. Chicago: American Occupational Medical Association; 1987.
- Felton JS. *Occupational Medical Management: A Guide to the Organization and Operation of In-Plant OHSs*. Boston: OEM Health Information; 1989.
- Guidotti TL. The general internist and occupational medicine. *J Gen Intern Med*. 1986;1:201–202.
- Guidotti TL, Cowell JWF, Jamieson GG. *OHSs: A Practical Approach*. Chicago: American Medical Association, 1989.
- Murray R, Schilling RSF. Functions of an OHS. In: *Occupational Health Practice*. London: Butterworths; 1981: 145–158.
- Newkirk WL, Jones LD. *OHSs: A Guide to Program Planning and Management*. Chicago: American Hospital Publishing; 1989.
- Rest K, Ed. How to begin an occupational health program—administrative and ethical issues (entire issue). *Semin Occup Med*. 1986;1(1):1–96.
- Smith I. *Occupational Health: A Practical Guide for Managers*. London: Taylor and Francis; 2007.

*This page intentionally left blank*

# **21 WORKERS' COMPENSATION**

Workers' compensation is a cornerstone of occupational medicine. The occupational and environmental medicine (OEM) physician should therefore invest considerable time and effort in understanding the principles of workers' compensation and learn as much about the system in which one practices as one possibly can. (There is no counterpart in environmental medicine.) At the outset, it should be noted that the correct form is "workers' compensation" (plural possessive), not "worker's compensation" (singular possessive), and that "workman's" or "workmen's" compensation is an obsolete, historical term that survives today in the names and texts of legislation passed in earlier times, when all workers were assumed to be men.

Eligibility for the insurance benefit is based on an injury or illness that "arises from work" or that "arises out of and in the course of employment," in the most common language. Causation and relationship to work is therefore an important, if implicit, part of every workers' compensation case. Issues of causation become complicated in occupational disease cases and require a systematic approach to evidence, aspects of which are discussed in Chapters 3 and 23.

Workers' compensation is a social insurance system, one of the earliest. It provides injured workers with two essential services: income

replacement and payment of medical and rehabilitation expenses. Workers' compensation began in Europe in the mid-nineteenth century and was first put into practice in Germany in 1884 in its modern form. It began as a great reform. Workers' compensation pioneered no-fault insurance, introduced alternative dispute resolution, created a counterweight to abusive power of employers, created incentives for safety and injury prevention, and saved countless families from sudden impoverishment. Workers' compensation was not introduced into North America until 1911, during the historic but eclectic movement for social reform known as the Progressive Era. Over time, inequities and abuses crept back into the system, which became rigid and parochial. The state-level organization of workers' compensation led to fragmentation, inconsistency between states, and heterogeneous policies. The system as a whole came to be seen as corrupt and unresponsive, dominated by lawyers who had reintroduced the litigiousness and delays that workers' compensation had originally sought to overcome. Over the last two decades, however, a great reform movement has reenergized workers' compensation. Now, workers' compensation systems are laboratories of experimentation and innovation in managed care.

Virtually all OEM physicians work at least in part within the workers' compensation system (colloquially and universally called "workers' comp"). Workers' compensation is one of three pillars of occupational health practice, the others being the occupational healthcare system and medicolegal practice, including independent medical examiner services. It is a parallel healthcare system sharing providers with the general healthcare system but funded and managed very differently. Workers' compensation is much older than managed care, arose from and was a response to the adversarial tort litigation system, and is not covered by much current legislation, such as provisions regarding confidentiality and access to information in general healthcare. These features have given workers' compensation different characteristics from other healthcare systems.

Today, workers' compensation reform can be read as a compromise between interests who would have preferred the status quo but realize

that it was unsustainable and those who would prefer a more thoroughgoing transformation but know that there is no political will for it. It has become the country's second largest healthcare system (about \$55 billion in 2003), coming after the general healthcare system (about \$1.7 trillion in 2005) and ahead of the military's Defense Health Program (about \$18 billion in 2005).

## **DESIGN AND PRINCIPLES**

Workers' compensation is a "no-fault" insurance plan designed to pay for medical care and to compensate the injured worker for wages lost during periods of temporary or permanent disability and rehabilitation due to injury or illness resulting from work. Critics of the no-fault aspect of workers' compensation point out that the absence of a means of identifying fault or a motive for doing so removes a check on the commitment of employers to run a safe workplace. Defenders point out that that is more properly the function of a regulatory agency and that the experience rating system is a means of penalizing employers that demonstrate poor performance.

An important principle of modern workers' compensation systems is that it is necessary to demonstrate that the injury arose out of work, but it is not necessary to find that an employer was at fault for a worker to receive benefits. Likewise, negligence or fault on the part of a worker does not disqualify him or her from entitlement to benefits. Causation analysis is essential in workers' compensation claims but is properly limited to establishing the relationship to work. This is usually obvious in an injury. The challenge of establishing causation in occupation disease is much greater, and at times the standard of persuasion that is needed requires much more detail on the circumstances of the injury than should be required to answer the essential question of whether the condition arose from work.

The general standard for adjudication in workers' compensation is the "weight of evidence," which derives from the tort system that gave birth to workers' compensation. Most workers' compensation acts tip the scale deliberately by specifying that when the balance of

probabilities is even, the benefit of the doubt should be given to the claimant. (Vermont does not.) Some states, such as California, and systems such as the Railroad Workers Act require the demonstration of a “substantial contribution” of work factors to the injury but do not require that the injury be completely the result of workplace hazards.

Workers’ compensation pays benefits for wage replacement and medical care as a primary responsibility. The objective is to compensate the injured worker for wages lost during periods of temporary or permanent disability and to pay for treatment and rehabilitation required to recover from injury or illness resulting from work. Retraining benefits are sometimes available. Incidental costs, such as transportation to and from treatment and for workers’ compensation board interviews or evaluations are sometimes provided. Benefits are also paid to survivors after fatalities.

Workers’ compensation introduced a new legal remedy in which the “industrial relationship” between employer and employee would be considered to be deeper than work in exchange for pay. The cost of healthcare for on-the-job injuries would be treated as an operating expense to be shared on an industry-wide basis. The worker would be guaranteed a fair and rapid adjudication of claims and sufficient compensation. (This, at least, was the ideal.)

Workers’ compensation is based on what is often called “an historic compromise.” The injured worker gave up the right to sue the employer and the employer gave up the right to reject the claim or to defend themselves in court. Under workers’ compensation legislation, the employer is required by law (except in Texas, where it is voluntary) to carry insurance to cover on-the-job injuries and compensation for disability. The employee, on the other hand, is not permitted to sue the employer for medical costs or compensation. The employee trades the uncertainty and delay of a court proceeding, which could possibly result in a large judgment, for prompt payment of predictably more limited benefits. The employer gives up the right to argue against a penalty but benefits by avoiding expensive litigation. This, at least, is how the system is supposed to work.

Workers' compensation is organized at the state and provincial level (with separate systems for federal employees in both the United States and Canada), and there is much variation in process and benefits. It was originally designed for industrial workers and so has large gaps in coverage, for example excluding agricultural workers in most states. It operates strictly on an insurance model, mostly through private commercial carriers, and is not a welfare or entitlement program. Workers' compensation is financed by a payroll premium levied on employers, not financed by taxes.

The essential services provided by workers' compensation are (1) income replacement for workers who cannot do their job because of injury or disease arising out of their work and (2) coverage for medical and rehabilitation expenses for the treatment of work-related injury and illness. Benefits are also provided to the dependents of workers in case of fatality. Individual workers' compensation carriers and boards may provide additional services, such as job retraining, but these two services are the foundation of the system. It follows logically, then, that the first decision in a workers' compensation claim is whether the injury or illness did in fact arise out of work. For most injuries, this is not in question. For diseases, including musculoskeletal disorders (which are treated as diseases in workers' compensation), the connection to work is not always obvious and may be disputed. As a consequence, there is always an appeals mechanism in the workers' compensation system.

Intensive management of care is another effective approach. Requesting regular progress reports, advising on management where appropriate, offering employee's facilities for rehabilitation, frequent visits to the occupational health department, and encouraging second opinions for major surgical interventions all tend to promote appropriate management of the case and provide opportunities for monitoring and evaluating progress. The procedure for evaluating fitness to work (see Chapter 18) can then be followed in order to return the worker to the job as soon as he or she is ready, minimizing unproductive time on compensation, and avoiding a return too early, when re-injury may occur.

The level of benefits varies greatly by jurisdiction. Within the United States, West Virginia, Maine, Montana, the Federal Employees system, Hawaii, Oklahoma, Washington, Alaska, and California pay the highest benefits per \$100 of covered wages, in that order. Comparisons may easily be misleading, however. Criteria and policies vary markedly, introducing great disparities in equity. States with heavy employment in high-risk industries, such as mining, may pay more. States with a large representation of more highly paid workers may pay less as a ratio to payroll.

The workers' compensation system is the sole legal recourse for injured workers that it covers. This means that a worker cannot sue, or be sued by, an employer. This is referred to as "exclusive remedy" and is written into workers' compensation acts. Although the exclusive remedy has been challenged in the courts, it has always held on appeal, if not in the initial ruling. Most of the challenges have been brought on the grounds that the "exclusive remedy" provision denies workers "due process" in having their cases heard. Also, lapses in the process itself, such as arbitrary actions or failures to comply with regulations on the part of the administrative judge or hearing officer, can themselves be grounds for legal action under some circumstances.

One opening for workers and employers is that the right to sue a third party is preserved to both. If a product used in the workplace is faulty and causes an injury during normal use or if the manufacturer or distributor did not effectively or honestly represent or communicate the hazard ("failure to warn"), there may be grounds for legal action. Likewise, if a chemical formulation causes an occupational health problem that should have been known and that potential problem was not communicated to the purchaser, grounds may exist for a lawsuit against the third party, usually the manufacturer or distributor. Third party suits have been heavily used in asbestos litigation. Ordinarily, judgments won against a third party are first applied to reimburse the employer or carrier for the value of benefits already paid to the worker, with the balance going to the worker directly. Third-party litigation is complicated, and expert legal counsel is advisable before proceeding with action under this or any other legal premise.

Workers compensation carriers also sue third parties and sometimes sue one another to obtain reimbursement for extraordinary claims. This is called “subrogation.” For example, the province of Alberta experienced a large number of claims for asbestos-associated disease in oil refinery workers who had also worked in Texas in jobs that also involved handling asbestos. The Workers’ Compensation Board of Alberta paid the benefits and then sued carriers in Texas for the share that would have been apportioned to the Texas companies and was able to recover millions of dollars.

In the United States, various state agencies, which may be industrial relations department, insurance commissions, or departments of banking or insurance, house offices of workers’ compensation regulation that mandate consistent adjudication processes, regulate insurance premiums, and provide an appeals process separate from the carrier, usually a tribunal system using an administrative law judge or a hearing officer. Appeals are considered to be a form of alternate dispute resolution rather than modeled after courtroom procedures and so the rules of evidence are relaxed. Certain “presumptions” are legislated in many states. Presumption means that certain claims are presumed to be meritorious unless proven otherwise because the disease in question is known to be associated with an occupation, requiring carriers to accept claims unless there is evidence to rebut the claim.

In Canada, workers’ compensation boards (WCBs) are provincial bodies created by law to operate the workers’ compensation system and workers’ compensation tribunals (in Québec, *la Commission des lésions professionnelles*) are autonomous agencies that exist to provide a fair appeals mechanism in resolving disputed claim. Workers’ compensation boards are empowered to set fees for compensation and medical reimbursement and guidelines for assessing disability. The “board” is the governing body of the WCB, variably appointed or selected but usually balanced among representatives of labor and industry as well as public servants. The WCB itself usually maintains its own staff of medical consultants as well as retaining authority to refer claimants to outside expert medical examiners and advisers when the claim warrants. Some WCBs also provide direct medical

and social services to injured workers, such as rehabilitation and vocational training, either in-house or by contracting out for delivery of these services.

Workers' compensation procedures have become exceedingly complex. Workers' compensation laws and rates are set by each state and province, with separate systems for federal employees and certain designated groups, such as longshoremen in the United States. There are seventy separate workers' compensation programs in North America, each with different procedures and levels of benefits. The system does not universally cover agricultural workers, self-employed persons, business owners, workplaces with less than a minimum number of employees in particular industries, domestic workers, or workers who are providing services on a contract basis as individuals. Volunteers, such as volunteer firefighters, are not covered. These omissions result in a substantial fraction of the workforce remaining unprotected and forced to rely on their general health insurance, welfare, or unemployment insurance as their safety nets.

Seamen in the U.S. Merchant Marine, Longshoremen (Longshore and Harbor Workers' Compensation Act), railroad workers (Railroad Workers Injury and Federal Employers Liability Act), federal employees (Federal Employees Compensation Act), maritime workers (Jones Act), coal miners (Black Lung Benefits Act), and federal vendors and federal-contractor energy workers in the nuclear industry (Energy Employees Occupational Illness Compensation Program, EEOICP) are covered by other systems in the United States, with special rules. The oldest of these systems predate the state workers' compensation acts. They are governed by federal acts in most cases because they involve workers in interstate trade or transportation. In general, they reset the burden of evidence such that injuries are compensable if there has been any substantial contribution to impairment, and they do not require the full weight of evidence to prove causation. The Longshoreman and Harbor Workers' Act of 1927, administered by the U.S. Department of Labor, created an insurance plan for shore-based maritime workers. EEOICP was initiated in 2001 to cover former employees of vendors and contractors at nuclear facilities for which the Atomic Energy Commission and then the Department of Energy

were responsible and who may be at increased risk for illness from chemical exposure. Administration of EEOICP was recently moved from the Department of Energy to the Department of Labor.

Physicians sometimes resist dealing with workers' compensation cases because the forms are tedious and the cases get complicated. Workers' compensation does require paperwork and administrative detail on the part of the clinician, but the system meets a serious social need and under current fee schedules usually reimburses the physician at a rate for the time spent that is better than many other cognitive (nonprocedural) healthcare services.

## **HISTORY**

Workers' compensation was the first social insurance plan to be introduced into North America, preceding unemployment insurance, as a cornerstone of protection for workers. It was a direct response to the social crisis that developed during the Industrial Revolution and the rising rates of injury and disability among workers in unsafe factories, mines, and workshops. Until workers' compensation was introduced, the injured worker had to sue his or her employer for compensation for an injury on the job.

Before workers' compensation, employees were considered under the law to have entered into an unwritten contract with the employer, in which they accepted all risk associated with their work as a condition of employment. Workers assumed legal responsibility for their own safety, even for measures beyond their control, and could sue employers only on grounds of criminal intent or criminal negligence, both of which were nearly impossible to prove.

During most of the nineteenth century, the United States and Canada were both developing countries with a chronic labor shortage (although not necessarily later and in the big cities) and dependent for growth on foreign direct investment (from London, in both cases). Law courts deliberately tried to protect the interest of employers in order to promote economic development, keep wages low, and make entrepreneurship attractive. The legal theories of the day stacked the deck against the worker.

Employers could readily defend themselves by invoking the prevailing legal doctrines or by confusing the picture of what happened. Originally, at the beginning of the eighteenth century, there was a doctrine in British common law (which applied to the colonies) called “the law of vicarious responsibility.” It held that the master was responsible for the acts of the servant and by extension the employer was responsible for the acts of the employee. However, this doctrine was largely blunted in the early 1800s by the “fellow servant” doctrine, which held that employers were not responsible for acts of their employees that resulted in injury to themselves or other employees. Before workers’ compensation, employees were considered under the law to have entered into an unwritten contract with the employer (“assumption of risk”), in which they accepted all risk associated with their work as a condition of employment. If a worker were injured by the action of another employee, the employer was not responsible (under the “fellow servant doctrine”). If any action by the injured worker could be construed to have contributed to the risk, it was defensible under the doctrine of “contributory negligence.” Sometimes juries sided with the injured worker anyway, but usually the legal defenses for employers were insurmountable.

By the late nineteenth century in Europe and later the United States, the Industrial Revolution had caused serious social discontent. The new industrial economy was based on hourly wage rates, production, and (especially in Europe) a labor surplus in cities (and, for several countries, their colonies) that caused competition among workers that kept wages low. Although wages in the United States and Canada were higher than in Europe during this period, workers were often divided, immigration caused short-term labor surpluses in the new industrial cities, and in the absence of unions workers could not act collectively to demand higher wages. In both North America and Europe technology threatened job security for workers who could be replaced by a machine and kept wages lower than they would have been otherwise, while at the same time keeping the prices of commodities low so that workers could live on them. In the United States, especially, technology and energy resources (water at

first, fossil fuels later) were used on a large scale to replace relatively scarce labor, somewhat offsetting the labor shortage, but these technological innovations were often very dangerous: open belts and gears, unguarded blades, inadequate ventilation, unprotected exposure to lead and acid fumes, and poor working conditions were the norm. Technology (initially borrowed from Britain) gave the rapidly growing American economy an early edge in industrial production, which it used to great advantage in the twentieth century. In Europe there had been revolutions in many countries around 1850, and there was a great fear of social and political instability. In the United States and Canada, there were periodic displays of rebellion and early labor unions were organizing, which raised similar fears. Labor insecurity was at the root of the discontent, although even at the time occupational health and safety was considered a side issue.

Employers hired workers “at will” and could fire them without cause. Social safety nets only existed through charity and mutual assistance pacts, which in the United States were usually based on ethnicity among immigrants and were considered suspiciously socialistic. The result was often a sense of desperation about finding and keeping a job. Work was often dangerous, fast-paced, and easy to lose if the worker could not keep up. When a worker was injured on the job, income for the family stopped, and if the injury resulted in permanent impairment, it could thrust the family from relative comfort into poverty almost immediately.

Occupational injuries, especially fatalities, were cheap in the nineteenth century because healthcare, which was not very good anyway, cost little and life expectancy was short. Employers occasionally recognized their responsibility and paid compensation or gave what they considered to be a charitable contribution to the family but usually paid nothing at all.

Workers’ compensation introduced a new legal remedy in which employer and employee were considered to have a special “industrial” relationship more complex than a simple contract for work in exchange for pay. It was based on a new theory that recognized that at least as an ideal both sides would have to give something up in

order to get something in return. The theory implicitly recognized that as owners of the facilities and means of production, employers are in control of the workplace and have a responsibility to workers who are injured or made ill in their employ. The outlines of a historic compromise took place, negotiated in part by early unions and workers' representatives but also by some employers who saw it as a way of reducing class conflict and introducing more stability into the workplace. The outline of the compromise was that workers would give up the right to sue their employers and employers would give up the right to refuse claims for injury. In return, the workers would receive a guarantee of reimbursement for lost wages and their medical expenses would be covered. The system would be run as an insurance business, using actuarial methods, which contained costs for the employers and avoided the risk of generous juries. Industrial sectors would have collective responsibility for their sector, because they would be insured on a group basis, which was thought to be an incentive to bring poorly performing employers in line. The system would be managed, or at least overseen, by government. Because it would cover all industrial workers and because fault or liability would not be at issue, the system would operate as no-fault insurance. Politically, the concept was irresistible because it received backing from both capital and labor and addressed a serious problem that was threatening social stability. The same unsettled social and economic conditions in the United States gave rise to the Progressive Era, a period of reform at the beginning of the twentieth century during which workers' compensation perfectly fit the political agenda.

Workers' compensation started in Europe in the mid-nineteenth century and was first put into practice in Germany in 1884 in something like its modern form (both workers and employers contributed to the fund, however). It was part of a package of reforms introduced by Chancellor Otto von Bismarck that was intended to quell political unrest and to head off revolution. (Other reforms included mandatory national healthcare insurance and the retirement pension.) Within a few years, other countries in Europe adopted similar systems, notably the Austro-Hungarian Empire, where many early inno-

vations occurred and where it became a big business. (Franz Kafka was employed in Prague by a workers' compensation carrier.) Workers' compensation was adopted in England in 1897 and operated until 1948, when the National Health Service was developed and a modified tort system for occupational injury and illness replaced it.

Workers' compensation was not introduced into North America until 1911, first in Wisconsin, Washington, and Ohio, when it arose as a reform movement during the Progressive Era of American politics. Most states adopted some form of workers' compensation on the English model (with less state management) by 1920 with the last being Hawaii in 1963. This history led to the present patchwork of programs and policies.

The German model (more centralization with state management) was introduced into Canada by way of Ontario by Chief Justice William Meredith in a highly influential report in 1913 that quickly shaped national policy and made workers' compensation policies more consistent across the country than in the United States. The essential principles articulated by Meredith were no-fault compensation, collective liability (within industrial sectors), security of payment (reliable and timely benefits delivered to workers), exclusive jurisdiction (keeping claims out of the courts), and an independent board to govern the system (to shield the system from political influence). The Canadian system thereafter developed along its own path with strong and large provincial "workers' compensation boards" organized as crown corporations, which serve as monopoly insurance carriers.

In the United States, the evolution of the system was different and emphasized the free and competitive insurance market in most states. The workers' compensation system worked in tandem with increased interest in occupational safety and had a stronger preventive component. Adjustments could be made in the "experience rating," which helped determine premiums paid by employers within a particular industry class, which created incentives to reduce the injury rate. However, as time went on, healthcare became more expensive and premiums became higher, the differential in payment that was governed by the experience rating became less significant as a cost. At

the same time, regulation ensured that the full increase in cost would not be transmitted to employers in disruptive jumps in premiums.

During the 1920s, workers' compensation expanded in coverage and programs and became the major social support system for working Americans. There was no other social insurance program at the time. Workers' compensation programs managed large pots of money and, unlike Canada, were not shielded from political influence.

During the Depression, workers' compensation was an incomplete but necessary safety net for many injured workers. However, perceptions of employment and safety changed. The dominant attitude was that workers were lucky to have jobs at all. Safety became more expensive to struggling employers than absorbing the cost in workers' compensation premiums. Federal efforts to impose requirements for uniform and effective occupational safety regulations on states were resisted. Workers' compensation systems faced a huge challenge in occupational disease claims, particular for silicosis, which imposed huge unfunded liabilities. In 1939, the U.S. federal government bailed out failing state workers' compensation systems nationally in order to keep the entire system from collapsing. One effect of this was to create more uniformity among programs than had existed before, especially in terms of standards and impairment ratings. However, the system remained fragmentary and controlled by state priorities and politics.

World War II, expenditures for which ended the Depression, led to more claims activity in workers' compensation, and protection of injured workers was perceived, at least to some extent, as patriotic because it supported the war effort. The postwar era was a period of general decline in efficiency and integrity of the system. The New Deal had introduced other important social insurance programs and employment benefits that made workers' compensation less economically important even while it expanded in assets and coverage. The country was preoccupied in the early 1950s with the debate over national health insurance, and there may have been a general assumption that such a plan would absorb workers' compensation, as had occurred in the United Kingdom, although the National Health Service was very different from the insurance plans under discussion

in the United States at the time. The workers' compensation system, marginalized and left to its own devices, became a political backwater, degenerating over time into special interest pleading, local fiefdoms, corruption, and political games played with money.

By 1972, the situation with state workers' compensation systems had reached the point of generating political pressure for reform at the federal level. In that year, a National Commission on State Workmens' Compensation Laws delivered a report that recommended mandatory standards for state systems. It did not explicitly recommend federalizing the workers' compensation system, but that would have been the logical next step if state programs did not comply with the recommendations. State programs did respond, however, and dramatically raised benefits to workers.

The 1980s saw the beginning of rapidly rising medical care costs and escalating premiums for employers. Caught in the middle because rates were regulated, the workers' compensation insurance industry lost money during this period. The pendulum swung back and led to a reduction in level of benefits, efforts to control costs, and the beginning of managed care models in workers' compensation. The industry became very profitable again.

The 1990s saw the beginning of major reform efforts, pioneered in states such as Florida and Massachusetts, and culminating in 2004 with thoroughgoing changes in California, the largest workers' compensation system.

What happened in California was a political phenomenon of the first order. In 2003, a sitting governor was recalled and Arnold Schwarzenegger, a movie actor, was elected to replace him in the recall election. Reforming workers' compensation became, uniquely, not just a matter of practical politics, as it always had been, but a demonstration of political will by the new governor, a means of establishing control, and obtaining buy-in from every important economic constituency (employers, injured workers, insurance carriers, providers, economic interests concerned with competitiveness, and firms considering a move into or out of California). With the possible exception of West Virginia, there is no other example in recent years

in which workers' compensation assumed such a central role in a state's political agenda.

The vehicle for workers' compensation reform in California was Senate Bill 899, which was passed in 2004. SB 899 allows workers up to \$10,000 in medical expenses paid for by employers while the claim is pending (before it is accepted or denied), requires that if injured workers wish to be treated by a physician other than the one selected by the employer they must do so within ten days and choose from a preapproved panel, allows closed healthcare provider networks for providing care, defines the qualifications of independent medical examiners (IMEs), caps access to rehabilitation services, replaces vocational rehabilitation benefits with a "supplemental job displacement benefit" if the injured worker could not return to his or her job, caps the duration of temporary disability to two years, creates incentives for returning workers with permanent disabilities to work, codifies the agreed medical examiner (AME) system, provides new avenues for dispute resolution, and allows system innovation under state supervision. In the world of workers' compensation, these are revolutionary changes. SB 899 also made evidence-based management the presumptive standard of care by imposing a new requirement for apportionment by cause in causation analysis (discussed later) and the adoption of evidence-based criteria for impairment assessment, treatment guidelines (using the *ACOEM Practice Guidelines*, which are discussed in Chapter 16), and utilization review.

Although the changes in workers' compensation have been invigorating and bold, the economic crisis of 2008–2009 initially threatens to upset everything. Whether it was at last achieving sustainability or not, the workers' compensation system in virtually every jurisdiction will face new and unanticipated challenges that will force even more radical changes.

The current or some future severe economic recession may force a crisis similar to 1939. It is not inconceivable that workers' compensation may require a federal bailout, which would create an opportunity to rationalize the system. Although for years there was no political will for a federal system of workers' compensation, the political landscape has abruptly changed, and it might become feasible again. Alternatively,

renewed movement toward a comprehensive healthcare insurance plan, although sidelined at the moment by the economic situation, might create circumstances similar to 1950 when workers' compensation could have been incorporated into a national healthcare insurance plan. The trajectory of current economic and political change guarantees that the issue of what to do with workers' compensation will be revisited in coming years.

## **CLAIMS MANAGEMENT**

A worker injured on the job is initially sent for medical care. In half of the states (such as Wisconsin) and the District of Columbia, the worker has his or her choice of which physician to see, but in other states (including Virginia and California) the employer makes the initial choice, although the worker can change physicians later if he or she desires. Some states (such as California) allow the worker to choose if they notify the employer before the injury of their physician of choice and increasingly often impose the restriction that the physician has to be on a list of approved providers. If the worker sees another physician, the cost of the encounter may not be covered by workers' compensation. In some cases, the insurance carrier may disqualify a particular physician or insist on another choice, particularly when there has been a pattern of suspected abuse or inadequate or ill-informed management of workers' compensation cases. The board or insurance carrier may also require periodic evaluation by another physician, sometimes called an "expert medical examiner" or "independent medical examiner," either on its own staff or an outside consultant. The purpose is to monitor progress and to have another set of eyes observing the treatment and recovery process.

The first visit to a physician for an occupational injury or illness does not require authorization from the workers' compensation carrier. For subsequent visits to be covered or for referral to a specialist to be reimbursed, care must be approved by the carrier in advance. The physician must file a "physician's first report" form within a certain time limit once an occupationally associated injury

or illness is recognized. This is a provision of law and is as legally binding as the requirement to notify the health department of a communicable disease. The workers' compensation carrier pays the physician a fixed fee set by the state for the report.

The physician's first report form initiates the claim. (Claims can also be initiated by employee's or employer's first reports, which are separate forms.) The first report automatically triggers a disability benefits review if the worker is off work more than a few days and establishes the injury as possibly work related.

The claim is initially reviewed by a case manager, who makes the decision to accept or reject the claim in the first instance. The case manager may accept or reject the claim on the basis of the first report, which is the means by which it is notified that a claim has been made, but seldom contests cases that are straightforward. The great majority of cases are not disputed.

The process of compensating the worker for wages lost because of disability is called indemnification. The intent is to compensate the injured worker for that proportion of future earnings lost due to the disability arising from the injury. Income benefits are triggered by the physician's or the employer's first report, which opens a file on the case. The benefits calculations are complex and are based on a formula that derives a disability rating, not on the level of impairment alone. Payment of temporary disability benefits begins after a brief waiting period (usually seven days) and continues until the worker has either recovered completely or has improved as much as he or she is likely to. In the former case, the payments cease. In the latter case, permanent partial disability payments are based on the disability rating that derives from residual impairment and other factors. (See Chapter 18.)

Until the worker can return to work, the carrier also pays for interim visits, consultations, and rehabilitation as long as the services are approved by the carrier in advance. If the disability is permanent, it is rated (see Chapter 18) and a weekly payment is made up to a maximum amount depending on the worker's capacity to work. Workers' compensation benefits cover only work-related disability. A person who is also impaired for some other reason must seek other sources of income maintenance for the non-work-related component.

After follow-up visits, written supplemental reports from the physician are provided to the carrier in order to monitor the course of recovery, control the continuation of benefits, and initiate payment for medical and rehabilitative services. Because progress reports by the physician are the means by which the carrier is kept informed of the case, it is critical that reports be submitted on a timely basis and that they be complete and accurate. The insurance carrier and the employer usually have the right to require an examination by a physician of their choice in order to monitor the progress of the patient-worker to determine when he or she is fit to return to work.

Complications of a work-related injury are treated as work related. Therefore misadventures in surgery, drug side effects, and medical complications are all covered by the medical treatment reimbursement provided by workers' compensation. However, incidental illness, second injuries that are not directly related to the initial injury and are not work related, or conditions that arise from personal health problems are not covered, even if they contribute to impairment and disability. These are assumed to be covered by other social insurance programs, such as Social Security. Some workers' compensation jurisdictions arrange offset payments coordinated with private insurance or social insurance programs, but most do not. No workers' compensation and few private insurance plans allow double collection of benefits for the same disability.

The final report terminates the case in the files of the carrier, establishes fitness to work (see Chapter 18), and establishes for the record any permanent disability that remains.

The workers' compensation carrier may choose to address questions of fact or opinions regarding diagnosis, impairment evaluation, and medical management by retaining an "independent medical examiner" (IME), whose function it is to resolve medical uncertainties. The IME may review the file and examine the injured worker but does not engage in treatment or assume care of the injured worker. The IME forms no physician-patient relationship with the injured worker; his or her sole responsibility is to render a disinterested and objective report on matters of fact and, if asked, to render

an opinion. In California, a variant of the IME system is used in which both claimants and the carrier agree on one physician, called the “agreed medical examiner” (AME) to perform the independent medical examination (which is also called an IME).

As in any benefits program, some level of abuse occurs in workers’ compensation. Usually, this takes the form of a worker claiming that an injury occurred on the job when it did not or of an employer concealing the frequency of injuries on the site by fraudulent or censored reporting. Sometimes, the degree of disability claimed is exaggerated by the worker or minimized by the employer. Outright malingering by workers appears to be relatively rare overall. More common is for a nonmeritorious claim to be put forward to see how far it will go by a claimant who may or may not be convinced of its merit but who has decided to let the system sort it out.

Mistaken claims, submitted in good faith by workers who feel that they have been injured but who are incorrect in this assumption, should not be considered abuse. They are part of the discrimination and evaluation problem, and it is the responsibility of the system to resolve them.

## **APPEALS**

The carrier may contest the claim on the grounds that it probably is not work related or that the level of impairment is not as great as described. A second level of review takes place within the workers’ compensation carrier, usually conducted by a staff or consultant physician or senior case manager. Either the worker or the carrier may appeal the second ruling or contest the amount of disability to a special board or tribunal that reviews such cases.

The tribunal is an autonomous body separate from the workers’ compensation carrier. It functions like an administrative law court, but the rules of evidence are relaxed and the process is less formal. The case is reviewed by an administrative judge or hearing officer and sometimes by a panel, with the support of consultants and dedicated staff, who are usually health professionals with prior experience

within the workers' compensation system. Hearings are often held, allowing oral presentation of arguments and examination of witnesses. Because of the quasi-juridical form of the hearing, most claimants feel the need for advocates to state their case and may hire lawyers or obtain representation from a union.

Decisions follow deliberation on the case and are based on the weight of evidence. The rationale behind the decision is explained in a report that is not unlike a court opinion but less formal. Unlike courts, however, workers' compensation tribunals are not bound by precedent. The reason is that claims are required, in most legislative acts, to be made on the basis of the facts of the individual case, not on other considerations or legal authority.

Disputed cases are usually those in which causation is not obvious. Workers' compensation systems were primarily designed to handle occupational injuries (which by definition in workers' compensation arise from a single event) and do not function very well in dealing with occupational diseases, repetitive strain injury (which is deemed a "disease" in workers' compensation because it does not arise from a single event and is therefore evaluated under the theory of "cumulative injury"), "mental mental" cases (a term of art meaning mental stress resulting in mental symptoms or impairment, which might otherwise be called psychogenic stress outcomes, as discussed in Chapter 13), unusual outcomes, or any condition that is also common in the general population or associated with smoking. In relatively rare cases, there may be a suspicion of malingering or the circumstances surrounding the injury may be complicated or suspicious (in other words, there is evidence that the worker may not have been injured on the job and may have sustained the injury elsewhere).

Cases in which there is no objective finding, such as low back pain, are especially difficult because symptoms reported by the patient have a strong subjective and emotional component, especially pain. Pain is subject to "symptom magnification" or "symptom amplification" (unconsciously increasing in perceived intensity due to anxiety or attention focused on the problem), "symptom exaggeration" (overinterpretation when the symptom is actually there), and

malingering (fraudulent representation). As a whole, the workers' compensation is seldom persuaded unless there is some evidence of an objective finding in a case. Because subjective complaints lend themselves more easily to false or self-serving representation than disorders with objective, documentable findings, case managers usually have a higher threshold for acceptance for back pain or chronic pain. While doubtless this helps to weed out nonmeritorious claims, it places a considerable and occasionally unattainable burden of proof on the claimant with a meritorious but subjective claim, such as regional complex pain syndrome.

Although workers' compensation carriers are reluctant to admit it, claims that are difficult to resolve or understand are usually rejected in the first instance. The appeals level is better suited to review and deal with them, but some claimants accept their initial rejection and never pursue the matter. Other claimants, particularly those who are aged, retired, or seriously disabled, may have little motivation to go through a long, drawn-out appeals process. Some may settle for other disability payments, through Social Security or Canada Pension (see Chapter 18) or not file a claim at all. This is thought to be the reason that the number of claims for asbestos-related disease is far less than would be predicted on the basis of the prevalence in working populations exposed to asbestos. Some workers' compensation carriers have special units for such difficult cases or cases that are likely to be disputed or may provide specialized technical support and consultation for the case managers in such cases.

Claims are normally reviewed first by a case manager in the role of an "adjudicator." This is an employee of the insurance carrier, which in Canada is the Workers' Compensation Board itself. The adjudicator is empowered to make a decision accepting or rejecting the claim on the basis of the evidence provided or to request further information. Most claims are easily adjudicated because they reflect obvious injuries and are not disputed by the employer. A small fraction, however, are ambiguous, incomplete, doubtful, unfamiliar to the adjudicator, or disputed by the employer. These are the problem cases that consume the time and attention of the WCB system as a whole.

When a claim is rejected or a disability rating is not to the satisfaction of an injured worker, an appeals mechanism is provided for review of the case. In the United States, this often takes the form of an administrative hearing, following a courtroom procedure. Often, however, the appeals hearings are rather informal and involve a minimum of procedural rules. The appeal may involve review of the documents only or an appearance before an administrative tribunal. Often, legal counsel is present, although advocates for the claimant do not have to be lawyers and some WCBs provide advocates on behalf of the claimants on a system much like that of public defenders in the courtroom. Decisions by this tribunal are usually binding but may be reviewed at a higher level on a discretionary basis, often by appearance before the board itself.

Two problems arise frequently in disputed cases. The first occurs when the workers' compensation carrier and a personal health insurance carrier both dispute a claim. This leaves the claimant without compensation and, after the workers' compensation appeals mechanism is exhausted, often leads to a lawsuit against the insurance company. The other is when temporary disability benefits expire but the workers' compensation carrier denies responsibility for permanent disability. The claimant must then apply for long-term disability (LTD) and is at risk that the claim will be denied. The transition from short-term disability to LTD is often a time when claims are re-evaluated and further opinions are sought.

## **DISABILITY EVALUATION**

Disability evaluation, in general, is a measure of how permanent impairment affects the capacity to participate in activities of life and work. Within workers' compensation, the grounds for evaluating impairment in workers' compensation are narrow and tied closely to employability. Disability based on the impairment is driven primarily by projections of the future job market for the worker and is prepared by actuaries, as it would be in other insurance.

On the basis of the information obtained, the adjudicator must make a judgment on whether to accept the claim and, if accepted, at what rating of disability. A disability rating is a complex, weighted percentage figure that relates the degree of functional impairment (as determined medically, as discussed in Chapter 18) and employability given the worker's level of education and training and the local or regional market for such skills. A disability rating may be temporary or permanent. Benefit payments for temporary disability are a set fraction of the worker's earnings before the injury (such as 80 percent of pre-injury income), which takes into account that expenses directly related to work, such as transportation, are reduced during the period of disability. Permanent disability causes most of the problems because it is the basis for either long-term payment or a lump sum reflecting diminished employment capability for the rest of the claimant's life.

Chapter 18 discusses impairment evaluation and the relationship between impairment and disability. Disability evaluation takes into account three basic elements:

- The impairment model (A): Impairment assessment and how it relates to disability evaluation is discussed in Chapter 18. Impairment in a given case may be evaluated against guidelines (such as the *AMA Guides to the Evaluation of Permanent Impairment*), schedules (awarding a fixed amount or assuming total disability in a particular condition), or job requirements.
- The wage loss formula (B): This element calculates the actual loss of wages as a direct result of the injury.
- Loss of earnings capacity (C): This model projects future loss of earnings as a result of the impairment (A), based on actuarial models.

Thus, the formula for rating disability is the sum of B and C (which is a function of A). Because disability evaluations start from the impairment evaluation, the level of impairment is often a central issue that requires clarification by an IME.

## **INSURANCE OPERATIONS**

Workers' compensation insurance carriers are insurance companies that write policies for employers for workers' compensation coverage. In the United States, they are usually private. The industry is highly concentrated with only a few companies writing the majority of policies nationwide in the United States (CAN, Fireman's Fund, Liberty Mutual, The Hartford, and Travelers). In twenty states, as in California, there are special state-administered insurance pools for companies that are not acceptable to private carriers, as an insurer of "last resort"; their premiums are expensive. In twenty-four states and the District of Columbia, there are only private insurance carriers. Self-insurance is allowed for large private firms in all but two states and is usually done by capitalizing a fund set up with a contractor to manage claims. In all Canadian provinces and in five American states (North Dakota, Ohio, Washington, West Virginia, and Wyoming) and two dependencies (Puerto Rico and the U.S. Virgin Islands), there are single-payer, publicly administered monopoly workers' compensation boards. The Federal Employees Compensation Act in the United States and the Government Workers' Compensation Act Labour Program in Canada are publicly funded through taxes. In British Columbia, uniquely, the Workers' Compensation Board is also the regulatory agency for occupational health.

Participation in workers' compensation is required by law for employers that qualify, except in Texas, where it is voluntary. Rates are assessed on the basis of the total payroll of the employer. The employer pays an annual assessment to the carrier for coverage, based on a standard rate to cover the frequency and amount of claims in the employer's industry, modified according to the number of occupational injury claims actually experienced by that employer in the years immediately preceding. The "experience rating" of an employer, as it is called, represents a means of providing an incentive to employers to improve their performance as well as protecting conscientious employers in the industry from the poor performance of their competitors.

Workers' compensation carriers have a variety of funds and reserve mechanisms to cover special situations. Reserves are often set aside for special categories of claims that are likely to result in expensive benefits. Asbestos-related disease is one such category.

"Second injury funds" are special funds or pools set aside under common or government regulatory management to cover claims for exacerbation by a subsequent injury after the first has left a worker with a disability. They reimburse the employer's workers' compensation carrier for the benefits for the differential in impairment between the initial and the later injuries and are not counted against the record of the employer in the experience rating. The reason they exist is to remove barriers to hiring disabled workers, one of which is that employers are often reluctant to hire workers with previous work-related injury on the grounds that they are more likely to injure themselves or are unsafe workers. Second injury funds relieve the employer and the carrier of financial risk for new impairment that arises because the second injury does not count against the employer's insurance rating.

"Unfunded liabilities" are projected future claims that may occur for which the carrier has not set aside funds. Some potential liabilities are unfunded because they may not occur and the carrier is assuming a risk. However, often unfunded liabilities were not appreciated or recognized, and the carrier is financially unprepared. Asbestos-related claims, for example, were a surprise to some carriers, who did not set aside enough in reserves to cover the anticipated costs.

Costs tend to be shifted between workers' compensation and unemployment insurance (UI), personal health insurance, long-term disability insurance (LTD) or Social Security, and welfare benefits. Workers who are injured off the job are not entitled to workers' compensation benefits but may apply for personal health insurance benefits and, if permanently disabled, for LTD. Many costs of occupational disease, for example, are never recognized or claimed, particularly when the worker is retired, and so are hidden in the costs of Medicare or Social Security Disability Insurance. Long-term care insurance and workers' compensation should be complementary, but

too often the claimant falls between the cracks if both of them disallow the claim. There is a strong incentive to file a claim for workers' compensation even when the claim is tenuous at best because the benefits are longer term, no copayment or deductible is charged to the patient, and there may be no or fewer restrictions regarding subsequent employment. Workers whose claims under workers' compensation are denied but who feel that they cannot work often apply for UI and for LTD (if they are eligible) and subsequently, when eligibility expires, welfare or Social Security Disability Insurance if they have no other source of income.

There are several ways within the complexities of the workers' compensation system that employers can minimize escalating workers' compensation system assessments. The best way, of course, is to improve performance by reducing the frequency of work-related injuries and illnesses by controlling hazards and training employees. A bad way is to fail to report relatively minor injuries, a distortion of the system which keeps the workers' compensation system, the regulatory system, occupational health and safety regulatory agencies, and the employer in the dark about a potentially serious problem and a violation of the law. Underreporting of minor injuries (in workers' compensation and in the OSHA 300 log) is very common. The worst way, however, is to fire or threaten to fire employees that are injured or report injuries. This is a reprehensible practice and collusion with it violates the ethics of occupational medicine. Unfortunately it is a common tactic of unscrupulous employers.

## **APPORTIONMENT**

In California in 2004, OEM and other physicians were required for the first time to apportion causation in workers' compensation reports. This was a truly revolutionary requirement. It reversed conventional thinking in workers' compensation, which was that apportionment may be applied to harm, impairment, or disability but never cause. However, it depends fundamentally on the judgment of clinicians, not on verifiable science.

Apportionment by cause is, in essence, an attempt to assess what percentage of a given level of impairment was caused by one causal factor compared to another. In principle, this would be very much like apportioning impairment, which workers' compensation carriers do routinely, except that the underlying cause of the disorder would be considered rather than how much impairment a second injury caused or a work-related injury on top of a pre-existing nonoccupational impairment. Apportionment of "harm" or "fault" is commonly used in tort law after a judgment to determine the share of damages that must be paid by each defendant. However, workers' compensation is a no-fault insurance plan, so apportionment of damages or by fault cannot be used.

The benefits of apportionment, if it could be done accurately and usefully, are obvious: adjudication would be simpler (especially if formulae could be worked out for relative proportions under certain assumptions), adjudication would be fairer to employers and some injured workers, financial resources would not be diverted away from injured workers and toward compensating for nonoccupational disease, incentives might be created for workers to take responsibility for their own health, the burden of disease would be more fairly shared among workers' compensation carriers and health insurers, and the relative contribution to disability benefits for permanent impairment could be divided up and allocated to different payers, such as pension plans and private long-term disability insurance.

Apportionment by cause is only useful across a narrow range of the total spectrum of causation. A particular cause may, conceptually, be apportioned as a sole cause, a sufficient cause, a substantial contributing factor, or a barely substantial but contributing factor, or not a significant factor. If an occupational cause is responsible for more than 50 percent of causation, then it would normally be treated as the sole cause. If causation is less than some nominal amount that represents a substantial contribution, say 10 percent, then it is not really significant in terms of its main effect, and the real question is whether it was the straw that broke the camel's back. In other words, whether "but for" that small contribution the injury or disease would have

occurred. Apportionment by cause therefore only operates meaningfully over less than half of the spectrum of causation.

Apportionment should not be confused with attribution, as the term is used in epidemiology (see Chapter 2). Attribution is a statistical approach, which applies only to populations. The attributable risk is the risk, and attributable risk fraction the proportion, of cases of a particular injury or disease that can be attributed to a particular risk factor. However, these estimates do not apply to individuals. Unless the individual belongs to a well-defined subgroup (in which case the confidence intervals are usually wide), the population risk estimate in an epidemiological study only represents a best estimate of what the risk would be for members of the group, chosen at random. Workers' compensation acts require that adjudication be based on the individual's circumstances, not on generalities. Furthermore, estimates of attributable risk fraction (the percentage of a disease associated with a certain risk or putatively causal factor) are hardly ever available for many risk factors in a single, relevant study. Picking and choosing estimates from different studies and then comparing them is methodologically suspect.

These problems essentially preclude the use of epidemiologic findings to make fine distinctions and careful calculations of apportionment by cause. Epidemiological findings provide best estimates and rough guides and cannot go much beyond an imprecise ranking of causes. (This discussion may be considered heresy among epidemiologists.)

On the other hand, epidemiological findings inform apportionment by providing insight in a general way and providing guidance on the magnitude of risk in the past. Knowledge of past risk for groups does help the physician in thinking about comparative risks for an individual but in an indirect way through individualized risk estimates.

The individual approach to apportionment is to take into account all the factors relevant to the individual claimant and to compare them with respect to their known relative strength (as reflected in epidemiological studies) and how much influence they are likely to

have had in the individual case. This approach can be done in two ways by (1) attempting to isolate the best applicable epidemiological risk estimate by looking at subgroups that most closely match the claimant's personal characteristics or (2) identifying personal risk factors and building a case for or against them being sufficiently substantial in magnitude to influence the outcome.

Building a risk profile from individual characteristics of the claimant is more difficult. The most important personal characteristic is usually age, especially for cancer and heart disease, and age-specific risk estimates are available for most common diseases. The age-specific risks associated with cigarette smoking are also available for many common diseases. Risk estimates associated with occupational exposures may be available from relevant epidemiological studies. One must then factor in family history, lifestyle, and other pertinent characteristics, for which risk estimates are not knowable. In practice, this means a judgment call in adjusting risk up or down in the individual case and in comparing relative contributions to the outcome.

The result is an uncertain mix of defensible quantitative risk estimates and subjective judgments with respect to risk patterns. Clinicians may be better able to adjust a risk estimate up or down compared to a known epidemiologically derived risk than they are in guessing the absolute risk with a "knowledge peg" (a term in cognitive psychology that refers to a prompt that starts speculation from a point of relative certainty). This does not mean guessing. It means applying the judgment of a clinician to a "post hoc" probability that takes into account prior knowledge and risk levels in the community but is difficult to calculate and communicate.

Although it is attractive to place greater emphasis on quantitative risk estimates, there is a good reason not to. That is because the application of conventional statistics to events in the past is not correct. (See Chapter 2 for the same argument as applied to cluster analysis.) Conventional, "frequentist" statistics are designed to describe events in the past and to predict the probability of events in the future (assuming no change in underlying factors). They are not designed to predict probabilities in the past. In other words, they do not accurately

describe the probability that a particular outcome would have happened in the past, from the point of view of an observer before the event occurred. This is the domain of Bayesian statistics, a methodology developed for clinical studies and other situations in which one is looking backward to see how likely it was that certain things happened. Bayesian statistics are much more complicated than frequentist statistics and are almost never used in medicolegal or workers' compensation.

Because of all these drawbacks, a more qualitative or subjective individualized risk apportionment may paradoxically be a more reasonable approach to the problem than a misleadingly sophisticated analysis using frequentist (conventional) statistics. In practice, apportionment by cause is inexact and often difficult. These limitations are acceptable, however, if they provided at least a rough and defensible guide for apportionment, as crude as halves, quarters, and bits rather than unattainable accuracy to the second decimal.

## **OCCUPATIONAL DISEASE**

In workers' compensation terminology, the term "disease" is applied to conditions that do not arise from a single event. Conditions such as noise-induced hearing loss, repetitive strain injury, chronic low back pain that did not start with a defined event, and other chronic disorders are classified as "diseases" rather than injuries because they are not associated with a single event, whether or not they involve the musculoskeletal system and whether or not they involve physical hazards and the release of energy (as discussed in Chapters 9 and 16). Occupational diseases present a particular challenge for workers' compensation.

Occupational disease cases are particularly difficult to manage in the workers' compensation system. They are greatly underreported and underrepresented in claims. As noted, there may be little motivation for retired workers, especially, to pursue a claim knowing that it will be disputed and that the benefits may not be great. There is no incentive on the part of carriers to go out and find cases. Occupational

disease may be difficult or even impossible to prove, especially if the latency period is long, there is no exposure data, and the disease could arise from factors outside of work.

The workers' compensation system developed with the implicit assumption that all or most occupational injuries could be recognized. When a visible injury occurs at work, the system tends to function well. When diagnostic criteria, subjective complaints, and chronic illness come into the picture, the system has much more difficulty dealing with the claim. This differential response to injury and disease has shaped the system so much that it can fairly be described as a mechanism oriented to the injury incident and unresponsive to the health status of the worker.

Three mechanisms have emerged as a means of dealing with the problem: the concept of cumulative injury or trauma, schedules of designated work-related diseases, and rebuttable presumptions.

Some chronic diseases and health conditions that cannot be attributed to just one event or the responsibility of one employer are evaluated under the legal doctrine of "cumulative injury," which assumes that the condition arose as the result of countless tiny injuries on the job that collectively added up to a single impairment. The theory of cumulative injury is applied to both noise-induced hearing loss and repetitive strain injury and assumes that the ultimate outcome arose from numerous small, even infinitesimal injuries. The concept of cumulative injury approaches this problem by assuming that the injury is the cumulative result of numerous small, discrete injuries of roughly equal magnitude that occurred at every opportunity for exposure. The doctrine sidesteps the need to demonstrate a single incident of injury, which is deeply embedded in the initial claims evaluation process. It allows apportionment of responsibility by time among various employers if the worker has had a mobile employment history. The doctrine allows the board to sidestep the difficult issue of apportioning causation and instead to apportion responsibility for its share of the cumulative effect to employers on the basis of the duration of exposure that occurred in each workplace, on the theory that each exposure opportunity provided a setting for injury.

For example, impairment from noise-induced hearing loss can be apportioned based on the relative loudness (if data are available) and duration of assignment to various loud workplaces under different employers. This makes it possible to assess the relative proportion of benefits between carriers insuring different employers. This concept works best for noise-induced hearing loss, in which it approximates the real situation, and less well for repetitive strain injuries, in which the relative timing, ergonomic issues, and specific circumstances play a greater role.

Slowly developing disorders that do not fit the cumulative event doctrine pose greater problems for workers' compensation. The injury is not immediately apparent and often gets worse with continued exposure, in the case of respiratory disorders. An individual may change jobs numerous times before the disorder becomes clinically apparent. In the case of long-term, developed health effects, such as cancer, the worker may even have retired before the disorder is recognized. There are usually technical and logistical problems to demonstrating causation. For example, for cases of occupational cancer, latency periods may be variable in length, recent exposures are relatively less important, and exposure data on earlier exposures are almost never available. As well, many occupational diseases, such as lung cancer, have nonoccupational causes such as smoking, and are known to be occupational only because they are statistically elevated in frequency compared to the general population.

There are two related approaches to the problem of occupational disease that can be used for disorders that do not fit the doctrine of cumulative injury. One is the "schedule," or designated list, and the other is the rebuttable presumption. "Presumption" refers to the policy that certain types of claims should normally be accepted because the disease risk is presumed.

A schedule is a list of designated compensable work-related diseases and is a form of presumption. They normally include disorders that are well established to be occupational in origin. The diseases are usually relatively rare and closely associated with a certain

occupation. Examples include mesothelioma and asbestos exposure, acute myelogenous leukemia and benzene, and nasal cancer and woodworkers. Such lists are in common use in some North American jurisdictions, such as Ontario and British Columbia, and in many other countries, such as Germany. The Black Lung Benefits Act is the best example in the United States: The test is simply to prove that the claimant has the disease and is a coal miner because only coal miners get it. The evidentiary burden to establish a schedule is to document that a particular illness is work related and should be added to the list. Thereafter, the burden on the claimant is to demonstrate that the disorder is, indeed, the diagnosis in question. Thus, schedules tend to grow very slowly. A criticism of scheduling of compensable disorders is that it is usually very difficult to add disorders to the list until the scientific evidence is so complete that many workers will have foregone compensation for years and may well have died before their eligibility is accepted. This is most often the case for newly unrecognized disorders and was well illustrated by the initial reluctance to recognize asbestosis for what it was.

“Rebuttable presumption” is, simply, the policy that claims should be normally but not always accepted when, all other things being equal, a claim received from a worker in a certain occupation is more likely than not to have arisen out of work, whether or not it is possible to prove the association in the individual case. This is demonstrated epidemiologically when the risk for the group is at least doubled compared to a reference population, which satisfies the legal criterion of weight of evidence, or “as or more likely than not.” A doubling of risk is equivalent to even odds, which is a mathematical statement that the risk arising out of work is equal to the risk of all other factors and corresponds to the adjudication standard of the weight of evidence. An example is the widespread presumption for certain cancers in firefighters, which is based on rarity and elevated risk (kidney cancer) or assumptions made on risk that are burdensome to prove (lung cancer).

Presumptions are normally rebuttable, meaning that the adjudicating body may also examine evidence in the individual case that supports or calls into question the individual claimant's risk profile (such as the personal smoking history) and the relationship to work (such as duration of employment). From the standpoint of social policy, a policy of rebuttable presumptions accepts that some nonmeritorious cases will inevitably be compensated in the interest of the greater good of ensuring that all meritorious cases are compensated. Presumptions lift a substantial burden of documentation and advocacy from the claimant and expedite management of claims. They are often legislated and consciously made less burdensome and more accessible for certain groups as a matter of social policy, such as presumptions for public safety personnel who take an uncommon level of personal risk in their work.

## **REFORM**

At a time when it is showing signs of renewal and reinvention, the workers' compensation system in the United States is, at the time of this writing (2009), under severe financial stress due to the current economic crisis. It remains to be seen whether reforms will continue or accelerate, or whether the system as a whole will become unsustainable. (The Canadian system may respond differently.)

The issues most prominent today in discussions regarding the workers' compensation system include the following:

- Financial solvency: The economic recession is bound to affect workers' compensation dramatically. Historically, injury rates drop during a recession because of reduced business activity. However, claim rates may rise because workers who are about to be laid off often submit workers' compensation claims, hoping for an income supplement for a partial disability that was previously tolerable. Carriers are likely to be pushed out of business through a combination of rising claims, reduced investment income, and

chronic business losses. Until recently, workers' compensation was a profitable line of business. Since about 2000, however, workers' compensation carriers have lost money, especially in California, and several carriers have withdrawn from states in which they did not have a competitive advantage. This has reduced competition, which in turn tended to push premiums higher, but because the carriers are regulated, the upward pressure has not been passed on in full to employers. One business problem is that administrative expenses for workers' compensation policies are high compared to other forms of insurance.

- Managed care: As the California experience demonstrates, workers' compensation has emerged after a late start as a highly managed healthcare system. Provider panels, utilization review, practice guidelines, and electronic medical records are incorporated into workers' compensation case management, and carriers tend to be much more directive than in the past. Unfortunately, this does not necessarily lead to proactive integration and coordination of care. Much of the management of workers' compensation is a logistical issue of getting various providers to a common understanding, tracking referrals and opinions, monitoring progress, and setting goals. There remains a disconnect that includes the treating physician and other parts of the system that reflects fragmentation in the general healthcare system.
- Increasing severity: Over twenty years there has been a trend toward increasing severity of disability among cases. This, of course, drives higher costs.
- Timely and sufficient benefits: In theory, the system was designed to provide payment much faster than would be the case if the worker had to sue the employer. For routine injuries, payment is reasonably prompt in most cases. However, any claim that is unusual, disputed, or requires further documentation is at risk for delay in processing and may require appeal. This often delays payment of benefits by months.

- Benefits: Benefits paid to injured workers tend to be low compared to other forms of insurance. They are calculated as a percentage of wages (typically two-thirds) and capped, so that high earners do not receive a proportionally greater benefit. Some states index benefits to current wages in the state. None take into account out-of-pocket expenses, indirect costs, or social costs of disability.
- Fee schedules: Until recently, medical fees were usually compensated at or below reasonable and customary charges, but in recent years the fee schedule has risen, and changes in the medical care system seem to have slowed the rise in payments for general care. As a result, there is now little difference in some fee schedules between payment from workers' compensation and from other sources; workers' compensation payment is also generally more reliable than collections from some other sources. Some jurisdictions still lag behind, however, and the history of inadequate compensation has influenced the attitudes of physicians, particularly those who are older, toward the system as a whole.
- Prevention of permanent disability: From a human perspective, permanent disability reduces the person's capacity to participate in society and, with it, earning potential. Improving the outcome of treatment and recovery after injury benefits the injured worker, reduces the level of residual disability, and reduces costs. Two-thirds of payments for disability benefits are for permanent disability, and it is a truism in workers' compensation that a relatively small number of high-cost claims, particularly back pain, drive the majority of the benefit total.
- Integration: Because occupational diseases are often difficult to recognize and because it is human nature to attempt to get the highest benefit to which one may feel one is entitled, there is considerable trading and cross-subsidy among workers' compensation, personal health insurance, Social Security, and LTD. As these all represent insurance schemes, as opposed to welfare,

which is a transfer payment mechanism, the proposal has frequently been put forward that they should be unified into a single system. New Zealand did something along these lines in 1974 by creating a streamlined, unitary no-fault personal injury compensation system that treated work-related and non-work-related claims similarly (for example, automotive injuries, medical misadventures, unintentional injury, and occupational injuries are all treated the same). The system is widely considered successful in achieving equity for workers, controlling costs, and operating efficiently and simply, but it had a rocky beginning.

- Panels: The establishment of national or regional boards of experts to agree on causation and level of impairment by consensus might significantly improve the quality and consistency of adjudication. This is standard procedure in Germany.
- Institutes and centers: In Canada, government entities have been created that conduct research, investigate theories of causation, and document evidence for important adjudication decisions. The reports are publicly available and may be peer reviewed. Ontario has organized this function within the broader Workplace Safety and Insurance Board (WSIB). Québec has organized centers of excellence within the adjudicating body (*la Commission de la santé et sécurité du travail*, CSST), and supports a research institute (*Institut de recherche Robert Sauvé en santé et en sécurité du travail*, IRSST).
- Federalization: Federalization of the workers' compensation system in the United States is not being discussed currently, but the option is likely to return to the national agenda as a consequence of the current economic crisis, especially if there is a need for a bailout. It is difficult to imagine the political constellation that would force it, since states are likely to oppose such a move, but there has not been an economic challenge to the system of this magnitude since 1939. The experience of 1972 suggests that the widespread perception of system failure

was not enough to trigger federalization, at least in that political era. On the other hand, the return of healthcare insurance to the national agenda, combined with conditions imposed by the federal government for financial assistance, might result in a national system (perhaps resembling that for federal workers), absorption into a broader healthcare insurance system (as in the United Kingdom), or some alternative as yet unknown.

The tort system continues to have its defenders, who point out that for all the perceived abuse, the tort system provides a mechanism to determine liability and therefore works toward justice and a fairer level of compensation to individual cases. The workers' compensation system, on the other hand, was created in a simpler era in which the tort process was not working to the advantage of the claimant but has become highly bureaucratic and unresponsive. By this thinking, the tort system should be kept intact as the ultimate resort for the claimant, because in the absence of the threat of tort action, employers are far less likely to be cooperative. Employers in this era are thought to fear court action more than workers do because of the unpredictability in awards inherent in the tort action. This is very much a minority view.

In the future, fundamental reform might return to the geographical origins of workers' compensation and consider as a model the current German system (the *Berufsgenossenschaften*), in which workers' compensation insurance is organized regionally as a nonprofit public utility controlled by governing boards to which are appointed representatives of labor, industry, and the public. (This is already the case in Canada.) The German insurance bodies occupy a much more central role in the healthcare system and exert stronger incentives toward prevention.

In 1998, the American College of Occupational and Environmental Medicine published "The Eight Best Ideas for Workers' Compensation Reform." They are worth revisiting:

1. View workplace injuries and illnesses as evidence of prevention failure and use them to target safety and enforcement programs.
2. Require active linkages between injury and illness care services, prevention strategies, and disability reduction programs.
3. Make sure that job-related health decisions are made by health-care professionals with appropriate training and expertise.
4. Expect active participation by both employers and injured workers.
5. Begin management of job and life disruption as soon as disability begins.
6. Use evaluation and ratings systems based on objective, standardized methods as the basis for awards for both physical impairment and vocational disability.
7. Encourage workers' compensation managed care organizations to innovate; when provider choice is limited, require proof of quality.
8. Demand better and more standardized data and use them to guide medical care, to direct reforms, and to inform purchasers.

## RESOURCES

- Guidotti TL, Rose SG. *Science on the Witness Stand: Evaluating Scientific Evidence in Law, Adjudication, and Policy*. Beverley Farms, MA: OEM Press; 2001.
- Sall RE. *Strategies in Workers' Compensation*. Lanham, MD: Hamilton Books; 2004.
- Strunin L, Boden LI. The workers' compensation system: worker friend or foe? *Am J Ind Med*. 2004;45:338–345.

## NOTEWORTHY READINGS

- American College of Occupational and Environmental Medicine. ACOEM's eight best ideas for workers' compensation reforms. *J Occup Environ Med*. 1998;40(3):207–208.
- Dembe AE, Fox SE, Himmelstein J. *Improving Workers' Compensation Medical Care: A National Challenge*. Beverley Farms, MA: OEM Press; 2003.
- Guidotti TL. Considering apportionment by cause: its methods and limitations. *J Workers Comp*. 1998;7(4):55–71.

- Guidotti TL. The Big Bang? An eventful year in workers' compensation. *Annu Rev Pub Health.* 2005;27:153–166.
- Guidotti TL, Cowell JWF, eds. Worker's compensation. *Occup Med: State Art Rev.* 1998;13(2):241–459. [Occupational Medicine: State of the Art Reviews was a monograph series. This issue was devoted to workers' compensation and has many articles of continuing interest.]
- Neumark D. The workers' compensation crisis in California: A primer. California Economic Policy (Public Policy Institute of California), San Francisco, January 2005. Available at [www.ppic.org/content/pubs/EP\\_105DNEP.pdf](http://www.ppic.org/content/pubs/EP_105DNEP.pdf).
- Nuckols TK, Wunn BO, Lim Y-W, Shaw R, Mattke S, et al. *Evaluating Medical Treatment Guideline Sets for Injured Workers in California*. Santa Monica, CA: Institute for Civil Justice, Rand Corporation; 2004.
- Pace NM, Reville RT, Galway L, Geller AB, Hayden O, Hill LA, et al. *Improving Dispute Resolution for California's Injured Workers*. Santa Monica, CA: Institute for Civil Justice, Rand Corporation; 2003.
- Peele PB, Tollerud DJ. Managed care in workers' compensation plans. *Annu Rev Pub Health.* 2001;22:1–13.
- Williams CT, Rena VP, Burton JF Jr. 2004. *Workers' Compensation: Benefits, Coverage, and Costs*, 2002. Washington, DC: National Academy of Social Insurance; August 2004.

*This page intentionally left blank*

## **22 EMERGENCY MANAGEMENT**

The occupational and environmental medicine (OEM) physician is increasingly valuable to employers for his or her potential to contribute to the survival of the enterprise, not just its efficient operation. After several years of highly visible disasters, some employers, especially in critical industry sectors, have come to a new realization of the criticality of OEM functions to business continuity. Awareness is spreading in the corporate sector, stirred by recognition of the profound threat to continuity of operations and the survival of key personnel presented by natural disasters, major industrial incidents, and intentional assaults.

OEM physicians are now called upon to act in response to public and management concern over threats from intentional assaults (including terrorism), unintentional incidents that may result in mass casualties, and natural disasters. The occupational health system has similar, and in some cases more robust, capabilities than the general public health system. The scope of responsibility of OEM physicians and the demand for expertise and instant response has grew from dealing with workplace-specific hazards and injuries to being prepared to confront mass-casualty events and other large-scale threats to the workplace. Occupational physicians now have an opportunity to contribute as

expert advisors on emergency management, including naturally occurring disasters, emerging and reemerging infectious diseases, and security issues involving chemical, biological, and radiological threats.

Employers in critical industries and government agencies have become deeply concerned with continuity of operations and the security of their personnel. The imperatives of corporate security and homeland defense since 2001 have, in turn, invigorated and expanded the mission of OEM for disaster planning and emergency management.

At many enterprises, employers and their employees rely on occupational health providers for health information and care with a focus on the threats, hazards, and injuries unique to workplace organizations. To be effective, the OEM physician must work as a partner in a coordinated effort by an emergency management team that includes managers, technical personnel, and emergency responders.

## **HISTORY**

The role of the occupational physician in emergency management evolved from the historic involvement of occupational physicians in disaster planning, workforce protection, and prevention of disease and injury.

### **Roots**

Since the Industrial Revolution, practitioners of occupational medicine have had to treat victims of mass casualties and industrial misadventures. The “industrial surgeon” was sometimes called upon to manage the consequences of injuries on a mass scale, such as railroad accidents. This expanded the scope of occupational medicine practice and led to a role for the physician in “accident prevention,” which at the time did not distinguish between prevention of individual injuries and mass casualties. Later, particularly as corporate medical departments were established and expanded in the early twentieth century, the occupational physician became a recognized resource for disaster planning and emergency response.

During the Second World War, the occupational physician played a valuable and respected role in ensuring safe and secure operation of critical industries. Interest in disaster planning and emergency response declined after the war except in hazardous industries, especially after the Bhopal catastrophe in 1984, when thousands of local residents in India died after release of a toxic gas from a local chemical plant owned by an American company. The response to the Bhopal disaster in the United States was the Emergency Planning and Community Right to Know Act (1986), which focused on the community's right to know of local hazards and not on prevention, security, and mitigation. For the most part, the role of OEM physicians in emergency management was marginalized. Even so the American College of Occupational and Environmental Medicine (ACOEM) continued to provide opportunities for its members to learn practical skills in disaster planning and emergency management. Anticipating the future threat, in 2000 ACOEM was among the first medical specialty organizations to train its members in practical issues in bioterrorism and homeland security.

### **Events at the Turn of the Century**

The OEM physician's role in disaster planning and emergency management returned to center stage after the tragic events of "9/11" (September 11, in shorthand American usage, of 2001) and especially the anthrax assaults in its aftermath. The terrorist assault using hijacked airplanes in an attempt to destroy the Pentagon was a strike at military power, but the assault on the twin towers of the World Trade Center in New York was directed at the financial system of the United States and targeted employees working in the financial district, as did the earlier bombing there in 1993. Following the catastrophe, a huge relocation effort moved the employees and executives of financial institutions to temporary locations in New Jersey, where business continuity was quickly restored. OEM physicians at these financial institutions played a key role in managing health problems that arose, especially among older executives

who were not well. (This story has not been told but should be in due time.)

During the immediate response to the emergency, many workers did not use personal protective equipment and ignored safe work practices in their desperation to rescue victims. The Occupational Health and Safety Administration (OSHA), the National Institute for Occupational Safety and Health (NIOSH), and New York state officials were not supported in efforts to encourage the use of safe work practices and personal protection among responders, although OSHA brought equipment to the scene. The immediate result may have been a much larger number of injuries than would otherwise have occurred (5,222 incidents), including one fatality and many eye disorders, which probably impeded the rescue and recovery effort significantly. Several years later, a high rate of chronic respiratory symptoms has been documented among responders. OEM physicians at area academic institutions have played a critical role in documenting these health effects.

In the aftermath of 9/11, anthrax assaults—almost certainly perpetrated by a biological warfare expert with unknown motives—specifically targeted workers in the communications industry, both print journalism (in the first attack) and television (on the ABC studios in New York), as well as politicians and employees at the U.S. Capitol. There was an obvious connection between the workplace and the anthrax threat in issues involving exposure of postal employees, two of whom died. OEM physicians in Washington, D.C., handled much of the care of workers who had been or could have been exposed. Immediately following the tragedy of 9/11, an ACOEM task force produced a guide to the management of mental health issues among survivors of mass assaults, disseminated it to all members and posted it on the College Web site—all within four days. This achievement was unique and widely admired among medical specialty organizations.

The SARS outbreak of 2003 stimulated an organized response among several OEM physicians who formed a network to share information and recommendations for the protection of employees

under their responsibility. This network, guided by information received by companies operating in the affected region, especially an airline, helped corporate medical directors to decide on policies for travel, whether to suspend local operations, and measures for worker protection. The safety of health care workers was also a major issue during the epidemic, especially in Toronto. Documentation of the outbreaks in Toronto, Hong Kong, and Singapore has relied heavily on research by academic OEM physicians in those places.

In 2003, ACOEM developed the Occupational Health Advisory Committee (OH-AC) as a resource for coordinating responses, accessing management resources, and sharing information in times of crisis. The OH-AC was a part of a larger health care and public health sector coordinating council for the purpose of information sharing, analysis, and coordination (ISAC), organized at the request of the U.S. Department of Homeland Security. Subsequently, ACOEM developed and supported the Occupational Health and Disaster Expert Network (OHDEN) as a support tool for occupational health professionals faced with the challenge of emergency management. Its mission was to provide occupational health professionals with what they need when they need it in time of crisis. ACOEM and OHDEN briefly played a role in the aftermath of Hurricane Katrina in 2005, as described in the next subsection, facilitating early return to work. Unfortunately, the effort could not be sustained due to lack of funding.

The potential role of the OEM physician was also highlighted by several highly visible plant incidents during the decade (such as the Texas BP refinery explosion in 2004) and the civilian smallpox vaccination preparedness campaign in 2003, which predated the modern vaccine and was unsuccessful in meeting its objectives.

### **Lessons from Hurricanes Katrina and Rita**

Hurricanes Katrina and Rita, both in 2005, and their aftermath had a profound effect on perceptions of OEM and emergency preparedness.

Three distinct and important lessons were learned as a result of these natural disasters:

- Natural disasters are no less a threat to national security than intentional assaults
- In an emergency, employers may be on their own for extended periods of time
- Occupational health professionals can help mitigate an emergency

The first lesson that Katrina and Rita taught is that natural disasters and public health crises are as much threats to national security as intentional assaults. The effects of Katrina exceeded that of 9/11 in causing widespread destruction and dislocating American life. An entire region that played a vital role in the American economy, and still played a unique role in the country's culture, ground to a halt after its devastation by a storm that was not only predictable but had been predicted for years. A hurricane of that magnitude and in that location had, in fact, been the scenario for a training exercise for agencies in the region just the year before.

During Katrina and Rita, about 19 percent of the nation's oil refining capacity and 25 percent of its oil producing capacity became unavailable. The country temporarily lost 13 percent of its natural gas capacity. Together, the storms destroyed 113 offshore oil and gas platforms. The Port of New Orleans, the major cargo transportation hub of the southeast United States, was closed to operations. Consequences of such magnitude are beyond the reach of conventional terrorist acts (excepting possibly a nuclear threat) and equally or more disruptive to our society, economy, and ability to respond to subsequent threats.

The failure of the federal response in the days following Katrina compounded any shortcomings of the state and local governments. It forced all Americans, including employers, to realize that in a serious event, they are likely to be on their own for much longer than they might have supposed.

Obviously, in such a situation major employers would be well advised to protect themselves from the most likely disasters with redundant or private protective measures but could not reasonably be expected to sustain a catastrophic loss of a magnitude to destabilize an entire region. The rationale for business continuity is not only to keep the enterprise viable (and workers employed) but to carry out critical economic functions and to supply essential services. An example of this was the rapid response of Wal-Mart in supplying food and essential goods within the region struck by Katrina within a few days of the hurricane, far exceeding the capacity of federal and local emergency agencies.

The third lesson from Katrina and Rita is that occupational health and safety professionals can and did play a constructive role in emergency management, as did public health professionals generally. Rescuers took personal risks to save the stranded citizens of New Orleans. Less dramatically, public health agencies quickly identified and documented the risks of water contamination, warned of risks from carbon monoxide from portable generators, identified dermatitis and wound infections as major health risks, and identified outbreaks of norovirus-induced gastroenteritis. The role of public health professionals was widely, if insufficiently, recognized as essential and constructive. However, occupational health professionals, who play a public health role for the employed population, also played a constructive role, although it was largely invisible at the time and played-out mostly within the private sector or dispersed in communities.

Occupational health clinics and OEM physicians and nurses treated the injured, from wherever they came. Occupational health professionals returned critical personnel to work as soon as it was possible, to hasten economic recovery and rebuilding. Occupational Safety and Health Administration professionals warned against hazards in the floodwaters and the destroyed, abandoned houses, but were stymied because supplies for personal protection were nowhere to be found. The many stories of what occupational health professionals achieved in the aftermath of Katrina need to be collected and documented.

The Occupational Health and Disaster Expert Network (OHDEN) was a Web-based support system for occupational health professionals in emergency management developed in 2003 and supported by ACOEM. At the time of Katrina, OHDEN was still a prototype, with about two dozen participants. During the worst week, immediately following the hurricane, OHDEN briefly became operational, providing occupational health professionals in the region, their managers elsewhere in the country, and experienced consultants and advisors with a place to go for useful information, best practices, frequent updates, and shared information. Much of this information was actually used in the field, for example recommendations were downloaded, evaluated, and incorporated into one oil company's emergency corporate policy on worker protection for their refinery employees. Two weeks later, drawing on the OHDEN experience and the questions that arose, the same team developed a one-hour seminar by teleconference sponsored by ACOEM. On September 13, 2006, 200 occupational health professionals heard recommendations and had their questions answered on how best to return workers in the region to their jobs safely and how to monitor returning workers for possible physical and mental disorders. The ISAC/OH-AC and OHDEN experience during Katrina can be interpreted as "proof of concept," demonstrating that occupational health professionals can play a constructive role in planning, consequence management, and recovery.

A third phase in the recent evolution of emergency management occurred with the recognition of "distributed emergencies" as a different class of event. Epidemics initiated by intentional release of communicable biological agents (thought most likely to be smallpox, should it occur) had stimulated contingency planning for protecting civilian populations against emergencies that play out over time, as in the case of epidemics, rather than as catastrophic events. However, the initial response tended to be focused on setting up special-purpose monitoring and rapid response systems, rather than a broader systems approach and reinforcement of the existing but resource-starved public health system. SARS, as noted, gave immediacy to these con-

cerns and demonstrated that in the event, the first line of defense is the public health system and employers, most obviously in the healthcare system. By 2007, the risk of an pandemic of avian influenza was the most prominent issue for planners and involved many OEM physicians outside of healthcare institutions, particularly in establishing policies and contingency planning for their employers. As it happened, the pandemic that came, in 2009, was of Influenza A H1N1, however many of the same precautions and plans applied. The H1N1 event, which is still playing out at the time of this writing, demonstrated the value of sound occupational health practice, in this case respiratory protection, and also revealed new problems, such as unexpected resistance to vaccination based on irrational beliefs and distrust of authority, and the need to deal with circulating rumors, for example the false idea that N95 respirators were dangerous for pregnant women.

These emergencies demonstrated that the capacity of occupational health professionals to act decisively and effectively to mitigate the consequences of a disaster is not just theoretical. It is real and proven.

## **ROLE OF OCCUPATIONAL HEALTH SERVICES**

Employers face a quandary in preparing for emergencies. On a small scale, they would want to be able to manage an emergency within the organization, in order to confine the damage and keep it from being too disruptive to the continuity of business. “Business continuity” is a major concern not only for the employer’s benefit but because employers in critical industries many need to furnish essential goods and services for the community without interruption. However, there is a practical limit to what they can prepare for. As companies consider emergencies of increasing magnitude, employers must rely on external assistance, such as fire departments and local health care providers. Beyond a certain point, when an emergency becomes so great that it disrupts the community on which the employer would depend for support, planning for emergency response may cease to be viable. Sort of that catastrophic

point, however, there is much that an employer can do to mitigate the effects of an emergency, through preparation, planning, and properly strengthening the occupational health service.

An occupational health service cannot, alone, deal with a disaster. Any definition of disaster, and there are many, emphasizes that it is a sudden event with severe adverse consequences that outstrips the capacity of local resources and requires external assistance. However, a strong and well-trained occupational health service can significantly enhance the chances of successful management and survival of a disaster and protect business continuity by managing health issues and serving as liaison with public health agencies. Larger, on-site services may even have the capacity to assist in the community response.

Disasters have a recognized “cycle” with distinct phases:

- Pre-event (during which planning and preparation may be occurring)
- The “prodromal phase” (in which there may be a warning)
- Impact (in which the most important factor is resilience)
- Rescue (when emergency management is most critical)
- Recovery (which involves reconstruction)

Business continuity, especially in critical industries, may be important to resupply and sustain the community during rescue efforts and is essential to recovery, either because of critical goods (such as food) and services (such as transportation) or by providing the basis for financial recovery in the affected region through employment and restoring the circulation of money.

Many concepts familiar to physicians from emergency medicine are useful in emergency management, such as triage. However, “emergency management” is broader than emergency medical care and fundamentally involves planning, coordination, communication, and integration at the level of the organization, not the provision of medical services. The OEM physician may or may not render emer-

gency medical assistance in the response phase but should play a central role in the preparedness, recovery, and mitigation phases of emergency management in any organization in which they are placed.

## **Occupational Health Services**

Occupational health services represent a workplace health system with similar, and in some cases, more robust capabilities than the general public health system. In the wake of a series of terrorist actions and natural disasters since the fall of 2001, the demands on and expectations of occupational physicians to play a central role in disaster preparedness and response have grown dramatically. Their scope of responsibility and the requirements for their expertise and instant response grew from dealing with workplace-specific hazards and injuries to being prepared to confront mass-casualty weapons and other large-scale threats to the workplace.

To perform these duties effectively requires committed time for preparedness activities and a unit that is structured and whose providers are trained to play such a role in time of crisis. However, it is costly and inefficient for even large corporations to dedicate a full staff and support structure for the management of an event that may or may not materialize. This is why adaptation of the existing occupational health service makes sense for many employers, especially those in critical or hazardous industries. Perhaps most attractive to cost-conscious managers, investment in expanding the emergency management capacity within an occupational health service is not “lost” if an event never occurs. The same systems support and enhance the traditional occupational health services that industry and government employees require and may lead to cost savings, increased productivity, and reduced liability in their own right.

Employers already have in place a structure on which to build to protect their operations and personnel in the form of their occupational health services. Adaptation of the existing occupational health service makes sense for many employers, especially those in critical or hazardous industries. Expanding the mission of the occupational health

service builds new efficiencies into the emergency response system. The same resources used for tracking employees' health can be used for surveillance to detect potential disease outbreaks due to bioterrorism. The technology of hazard identification and measurement can be applied to detect chemical or radiation threats. The medical staff on duty primarily to monitor health and to provide timely clinical care to workers can provide surge capacity to the community in time of crisis. Clinical health services can be applied to keep key personnel on the job and safe, especially when they are moved to new locations or are operating under conditions of stress and extreme risk.

The usefulness of a trained, well-informed, prequalified medical resource for dealing with incidents on site is obvious. These may include, but are certainly not limited to, sending infectious material through the mail to company personnel (anthrax and the much more common anthrax threat, or anonymous "white powder"), the threat of company equipment (such as chemical storage facilities) used as instruments of assault, and managing the psychological consequences of an assault. The OEM physician, who is trained in hazard assessment, may also assume the responsibility of determining when a site is safe to reenter or a facility to be reopened.

An enterprise may be in a better position to control its liability and potential loss if it develops a flexible, effective emergency management capability within its occupational health services. The employer would also be able to show due diligence in anticipating and preparing for plausible threats. This could reduce its exposure for punitive awards or claims based on negligence or omission, if despite its best efforts something does occur. A company that is seen to be prepared is less likely to be accused after the fact of ignoring a foreseeable threat.

Less obvious, but equally valuable, is the role that occupational health services may play in managing the consequences of widespread disruption of business operations due to major threats and of protecting the business, the product, and the brand against catastrophe in cases in which a company's products, facilities, or operations are used to deliver a threat or as targets for terrorist activity. In time of crisis, the occupational health service may help get community

back on its feet by helping to keep an employer open or critical infrastructure functioning. Within the occupational health service, the OEM physician may be called upon to manage the corporate response to serious health-related issues, such as travel to areas in emerging infections are a risk, rapid investigation of suspicious outbreaks of disease or following exposure to potential hazards, and to determine when reentry and re-occupancy is possible in contaminated facilities, such as occurred after the decontamination of post office facilities contaminated with anthrax.

### **Role of the OEM Physician**

The OEM physician has a critical role to play in disaster preparedness and emergency management. The role includes, but is not limited to, protecting the workforce, preserving business continuity, preventing plant- or facility-specific incidents, planning for the mobilization of resources in the event of a community-wide catastrophe, vulnerability assessment, risk communication, and collaborating with community resources and authorities, such as public health departments and regulatory agencies.

The OEM physician is well prepared to work with management and technical personnel at the plant, enterprise, or corporate level in preparing for plausible incidents, planning for an effective response, identifying resources that will be required, and advising on their deployment.

The role of the OEM physician in emergency management is distinct from those of emergency medicine and emergency management personnel. The typical occupational physician is not a specialist in emergency medicine, an expert in emergency management and incident command, or a safety engineer, notwithstanding that many individual occupational physicians do have special expertise in these areas.

The role of the OEM physician in disaster preparedness is also distinct from those of safety engineering and risk managers. The occupational physician is generally well-prepared to work with other professionals, which may include occupational health nurses, engineers, industrial hygienists, emergency medicine specialists, or

emergency management professionals, on a team focused on the response to an incident. The skills brought to the table by OEM physicians are complementary and reflect the unique value of both public health and clinical training.

The OEM physician is well equipped to manage many aspects of disaster planning and emergency management because relevant knowledge and skills are already core competencies in occupational and environmental medicine, including:

- Knowledge of specific threats, including a broad range of chemical, biological, and physical hazards
- Knowledge of personal protection and other applied approaches to health protection and the skills to evaluate the adequacy of protection at the individual level
- A systematic approach to monitoring and protecting the health of populations, that is, people in groups
- A systematic approach to monitoring and protecting the health of workers and other persons at risk, that is, people as individuals
- Skills in managing behavioral factors associated with the workplace and stressful events
- Detailed knowledge of individual plants, working populations, communities, and resources within their areas of responsibility
- Managerial skills and the skills to effect change through policies and management of information
- Clinical skills and an understanding of appropriate utilization
- Working knowledge of regulations, regulatory compliance, and the structure of government agencies responsible for health protection at most relevant levels
- Experience in evaluating individuals for fitness to work, which may be applied in emergency situations
- Experience in evaluating workplaces for safety and health protection, which may be applied in emergency situations

- Expertise in risk management, including risk communication in an emergency
- Occupational health management for first responders

## **First Responders**

Details on the management of occupational health for first responders are beyond the scope of this book, but this subsection will outline some general principles. The term “first responder” has been extended (by the homeland security community) to include those who are first to receive the victims in the emergency room. First responders in a workplace situation are usually corporate security personnel who arrive on the scene before public safety personnel have time to get there. Many employers, particularly in the mining and oil and gas industries, have their own firefighters, rescue teams, and hazmat teams.

The overriding priority of first responders is to protect the victims and to secure the location. They rescue or otherwise protect others who are not able to save themselves. If the personal risk is acceptable, they have as a secondary objective that of protecting property from destruction or damage. To achieve these priorities, first responders allow themselves to be exposed to hazards that are unusual for anyone else in the community and that would not be tolerated in other occupations.

The occupations normally considered among first responders include the following:

- Police
- Fire fighters (often cross-trained as EMS and hazmat personnel)
- Emergency medical service (EMS) personnel
- Rescue personnel
- Volunteers (who should be appropriately supervised)

The occupational risks of first responders include the same hazards that threaten the victims. Because they often arrive before the site has

been secured, decontaminated, or thoroughly searched, first responders themselves face threatening situations on arrival. For example, the first arrivals at the site of a terrorist bombing must face the real possibility of a second explosive device intended for them. Even after the site is secured, first responders are confronted with events and circumstances outside the usual experience of human beings in their daily lives.

The occupational hazards experienced by first responders depend on their role at the scene. Physical hazards predominate, since first responders are often working in unsafe, unknown conditions. Police are more likely to experience intentional injury through violence, whether directed at them or as a consequence of controlling and restraining others. Firefighters are particularly subject to thermal stress and ergonomic stress. Hazmat workers may be adequately protected from chemical hazards by their protective suits and gear but experience thermal stress working with that same gear in hot weather conditions or near fires. Because of the many ways and frequency with which heat stress occurs as a problem in emergency situations, the provision of clean water and breaks for hydration is an important aspect of sustaining the overall response.

There are many physical dangers at scenes of explosions, earthquakes, or natural surface disasters. These include unstable structures, live wires, jagged surfaces and sharp edges, and fire hazard from ruptured gas lines. Injuries can be minimized by intensive training, job experience, strict pre-placement screening, and physical fitness, but personal protection and safe work procedures are also required. There is an understandable tendency to ignore personal safety in the rush to save another, but all too often the result is an avoidable casualty. An injured first responder on the scene not only becomes unavailable to do his or her job but imposes a burden on a rescue and first response system that needs to reserve its full effort for the victims.

Psychological adjustment factors are important in all public safety occupations. The nature of the work, involving extraordinary risk and experiences of danger beyond that is normal life, place first responders at risk for post-traumatic stress disorder, especially if they are involved in incidents in which victims cannot be saved or a team member dies.

Within the field there is an on-going debate over the effectiveness of “postincident debriefing,” a procedure in which responders talk in groups about what happened and their feelings about the events. This intervention used to be promoted extensively in public safety departments, but many responders disliked it. It is now much less often used. Recent research suggests that “reliving” traumatic events is highly stressful and that this intervention does not help. A strong support network and an invitation for those who want to talk about the past to do so in a safe environment may be superior.

Although each of the occupations that constitute first responders has its own set of hazards, risks, and traditions, first responders share several features in common including:

- An awareness of personal danger, often accompanied by coping mechanisms that may include denial;
- Long periods of relative quiet or routine interrupted abruptly by periods of intense activity often accompanied by psychological stress;
- Rigid codes of behavior and high expectations for performance, often accompanied by complicated job responsibilities and guidelines and high penalties for failure;
- A strong ethic of teamwork and comradery, always with a strong sense of mutual reliance and social penalties for letting down one’s co-workers; and
- A rigid hierarchy or “chain of command,” which is necessary in order to reduce uncertainty and to make sure that procedures are followed correctly.

These characteristics shape the culture of first responders and may make it difficult for them to “open up” psychologically. They also create a strong social support system that strengthens the responder community and makes it psychologically resilient.

## Services and Functions in Emergency Management

The most essential function of the OEM physician and other health professionals is planning. Planning for foreseeable industrial disasters can inform and refine the response to unforeseen threats, given that sophisticated disaster planning is a matter of identifying resources and contingencies for all threats, not deriving detailed plans for single-threat incidents. Planning must be done across the organization, with participation from all relevant management groups (including human resources, operations, maintenance, security, inventory, finance, legal affairs, and, of course, occupational health and safety). Planning must be coordinated with local public health agencies, hospitals, insurance carriers, and health care providers and may require involvement of suppliers, transportation carriers, and local utilities.

Planning for contingencies at a particular site should take into account the layout of the site, natural and constructed barriers, threats and hazards on the site, and routes for evacuation, means to secure the premises while preserving access for ambulances and first responders for operational response (for example, for staging rescue operations, triage, stabilizing casualties, decontamination, and “incident command” activities). Even locations without special hazards may benefit from such contingency planning in the event of an external threat. For example, the first anthrax assault was in the office of a newspaper, not normally a high-risk location. Decontamination may have to be continued at the hospital or a second location away from the industrial incident.

Traditionally, the corporate medical director usually assumed responsibility within the organization for planning the medical response to emergencies, identifying facilities and resources for dealing with serious injuries and mass casualties, and providing health protection for key personnel if required. Although outsourcing has reduced the direct involvement of occupational physicians in planning emergency management in many organizations, particularly in the service sector, this important function has not been replaced by external con-

sultants because it requires a practitioner with intimate knowledge of the operations, hazards, workforce, and policies of the organization.

The OEM physician can also add value to the management of catastrophic consequences in other ways. These include:

- Representing the employer to public safety and health agencies during the event (and cultivating relationships with them beforehand)
- Survival of key personnel in a catastrophic event
- Continuity of business following a catastrophic event
- Obtaining resources for assistance in a health-related emergency
- Surveillance of the workforce and the early detection of an outbreak
- Integration of emergency response with public health agencies
- Surge capacity in the event of a local event requiring mobilization of all available medical resources
- Vaccination programs and other protective measures
- Establishing on-site consequence management and mitigation programs
- Developing decontamination plans
- Creation effective personal protective equipment (PPE)
- Clearance for reentry to facilities, such as contaminated sites
- Liaison with “local emergency planning committees,” which are councils for emergency planning for chemical disasters, pre-hospital care providers, and local hospitals
- Continuing education and training on-site and in the community of indigenous risks inherent to the operation
- Access to MSDSs (material safety data sheets; see Chapter 14) for hazardous chemicals and other hazard information
- Advising public health agencies on particular hazards, their risk, and appropriate treatment (for example, hydrofluoric acid, which

is used in industrial settings but is not generally familiar to public health officials)

- Lead any after-action discussion to effect process and system improvement
- Triage, dealing with the “worried well,” persons with health concerns, and ruling out health problems so that attention can be focused on urgent issues

## **BUILDING CAPACITY**

The key to building capacity to manage emergencies within the occupational health service is to build an effective and efficient team. Teamwork comes from training and planning but also from regular personal contact and cooperation. A team that functions well in the complex duties of an occupational health services and that already knows the operations, workforce, and facilities is more likely to function well in an emergency than an outside provider—who may not be around in a crisis.

Training is the way in which effective teams are built. In emergency management, training places a heavy emphasis on simulations and on “table-top exercises,” which are more or less realistic scenarios that are worked through in real time, with participants playing various roles as responders. The usefulness of these exercises depends in large part on how realistic the scenario is and how the situation is made to change over time as it unfolds, in order to simulate unpredictable events. Training is critical to readiness and effective response, but it is not completely innocuous. Recent research suggests that people who role-play during simulations may experience some of the same psychological stress reactions as those who experience an actual event.

Acquiring the necessary expertise is obviously required. The occupational health staff may require special training to take on the additional functions, but this is not much of a stretch from current duties. County emergency managers are eager to share training opportuni-

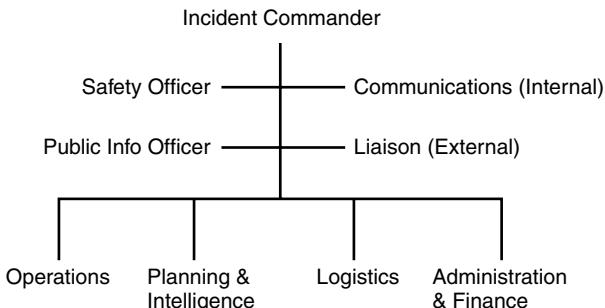
ties through grants and other programs within the public domain. The expense for preparedness may be justified by potential reductions in insurance premiums.

Establishing networks and agreements for mutual assistance may be critical. Here the occupational health staff can coordinate arrangements with local hospitals, specialist practitioners, public health agencies, and first responders in advance and maintain personal relationships required for smooth operation in the event of a crisis. The occupational health service may be able to provide surge capacity to the community in times of extreme need.

Certain routine functions can be anticipated and planned for. For example, if anthrax or some other threat is suspected in the mailroom, procedures (in the case of anthrax, quite simple) can be put in place in advance to protect employees, limit disruption, and rapidly evaluate evolving situations. This last function is particularly important to deter inevitable hoaxes and to prevent disruptions to business from ill-defined or unknown hazards. For example, the common scenario of an unknown “white powder” appearing on a loading dock or in an office can shut down operations for a day or more until a toxic substance is ruled out. Having the capacity on hand to show that it is harmless saves time and anxiety.

Likewise, dealing with panic among employees from rumors or incidental illness occurring at the worksite requires skill in rapid assessment and in risk communication (see Chapter 7) but can save an enterprise from devastating loss of confidence and the potential loss from employees who may refuse to come to work.

OEM physicians engaged in emergency management should be familiar with the “incident command system” (ICS), which is used by most public safety organizations, is being adopted by public health agencies, and is the conceptual basis for the U.S. National Emergency Management Plan. The ICS was developed in order to eliminate confusion in terminology, roles, and authority during emergencies. It is designed to be “interoperable” (operational regardless of the agency or setting), unambiguous, accountable, and scalable (it can be increased to the level of entire organizations or decreased to describe the different



**Figure 22.1.** The incident command system at its most basic.

tasks of a single person). The basic structure of an ICS is illustrated in Figure 22.1. When a public safety or health unit arrives on the scene or takes control of a situation, command authority is automatically given, all other things being equal, to the person with the most experience handling that type of problem, regardless of rank. The responders are divided into four line units, all under the direction of the incident commander: operations, planning and intelligence, logistics, and administration and finance. “Operations” deals with implementing the response, which in the case of a fire would be the unit responding, and in a broad and serious emergency may coordinate several public safety agencies (for example, police, fire, hazmat response, emergency medical services, and public health agencies). “Planning and intelligence” handles incoming information and plans the next step in the response, providing the incident commander with guidance on strategy and tactics. “Logistics” coordinates the movement of staff and supplies, to ensure that the right assets are available and deployed. “Administration and finance” keeps track of expenses and keeps a log of what is done, so that reimbursement for costs can be settled later. The incident commander is supported by four staff functions: one is responsible for maintaining internal communications among the responders (today this would often be an IT function); a safety officer who is tasked with protecting the health and safety of the responders; a liaison officer with government agencies, public safety units responding on request because of mutual assistance

agreements, and other stakeholders, such as local employers; and a public information officer who serves as sole spokesperson and the single source for external communications with media local residents, callers and outside inquiries.

## CHEMICAL HAZARDS

Among the worst plausible scenarios in industrial accident prevention and management are fires and explosions in chemical facilities. The worst chemical disaster in history from a fixed site was the famous incident in Bhopal, India, in 1984, which killed at least 3,800 people. In recent years, industrial communities have also needed to face the threat of intentional assaults using stationary sites as commandeered weapons (see section on terrorism). Most uncontrolled chemical releases occur during transportation incidents, such as derailed railroad tanker cars or trucking accidents.

“Chemical safety” is the general term for managing the hazards and risk associated with chemicals in order to protect workers and community residents from incidents of uncontrolled release or exposure. Chemical safety primarily means risk management, reducing the chance of an incident in the first place, reducing the release of chemicals if something does happen, and reducing the damage that the chemicals can do if they are released.

The increased concern over chemical safety has motivated changes to reduce the vulnerability of the entire supply and transportation system as well as increasing security for facilities on site such as stronger and safer tanks, greater control and monitoring of chemical production, and particularly substitution of more hazardous chemicals with less hazardous alternatives.

Moving toward increasing chemical safety and security is a strategy for “no regrets,” meaning that the investment in a safer system is not lost, even if nothing untoward occurs. Effective interventions change the vulnerability of the entire supply and transportation system as well as facilities on site. Historically, transportation systems have been even more vulnerable to incidents than plant sites and storage facilities.

When substitutions are made of less hazardous chemical substances, the supply chain and transportation system—the tankers, pipelines, and barges carrying the materials—also become safer.

Since 1985, with the initiation of an industry-wide program called Responsible Care®, the emphasis in chemical safety has been on anticipating and mitigating the risk of unintentional incidents and releases, particularly to neighboring communities and along transportation corridors. Chemical safety since 2001 has emphasized prevention of opportunistic acts of terrorism on site or the commandeering of hazardous materials for intentional assaults.

The primary concerns associated with particular chemicals are:

- Explosion
- Flammability
- Acute toxicity
- Exothermy (release of heat)
- Corrosiveness.

Responsible Care® requires participating companies to practice stewardship of chemicals throughout the product lifecycle. The essential elements of Responsible Care® are community participation, planning for emergencies, emissions reduction, occupational safety, and environmental sustainability. It is now worldwide.

Another program, mandated by the U.S. Environmental Protection Agency, requires community consultation on chemical facilities and the formation of a Local Emergency Planning Committee (LECP) where chemicals are manufactured, stored, or shipped. The LECP provides a forum for developing emergency response plans and disseminating information.

Chemical incidents have strict reporting requirements. They must be reported and may be investigated by the Chemical Safety and Hazard Investigation Board in the United States.

## PANDEMIC INFLUENZA AND DISTRIBUTED EMERGENCIES

Some emergency situations develop over time and not in one place; these have been called “distributed emergencies.” Global pandemics of infectious disease are the prototype for such emergencies, and their management provides a general model for dealing with emergencies that unfold more in time rather than space. These emergencies are not confined to infectious disease, however. The same model may apply to new drugs of abuse as they become popular in community after community, catastrophic changes in longevity (such as occurred in the central republics of the former Soviet Union after its dissolution), and famine. Infectious disease, however, is a present threat as well as clearer model for thinking about distributed emergencies.

At the time of writing (2009), the most credible immediate threat of widespread serious illness remains the H<sub>5</sub>N<sub>1</sub> strain of influenza, an avian disease that has already crossed the species barrier and has caused several human outbreaks. (The novel influenza A H<sub>1</sub>N<sub>1</sub> is the current pandemic, but mortality is not heavy.) The major risk is that a substantial number of people around the world will die or become incapacitated for a time, disrupting essential services. Employers in critical sectors, such as transportation, food production, water treatment, public safety, and especially health care, have to be concerned about how well they will manage when as many as 40 percent of workers may be incapacitated at the same time (the U.S. federal government’s scenario for planning).

The World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) are following the situation closely but depend on national and local health departments to monitor the progress of bird infections and to identify human outbreaks. The recent experience with SARS suggests that reporting is not always complete or timely.

The risk is that an outbreak of influenza will become a “pandemic” (worldwide epidemic) with the same or comparable virulence as the 1918 influenza pandemic, which caused millions of

deaths worldwide and was unusual in that the influenza pneumonia it caused was most severe in the young and fit. Influenza outbreaks do occur every few decades, although not of that severity, and the world is overdue. It is not clear what the characteristics of a future pandemic might be or what specific antigenic determinants the virus will have that will affect vaccine development.

In the event of a pandemic, WHO, CDC, and Canadian agencies will provide extensive guidance on their Web sites, as they do now for pandemic preparedness. The WHO Web site ([http://www.who.int/csr/resources/publications/influenza/WHO\\_CDS\\_CSR\\_GIP\\_2005\\_4/en/](http://www.who.int/csr/resources/publications/influenza/WHO_CDS_CSR_GIP_2005_4/en/)) is at the level of national planning. The U.S. government has an interagency Web site that relies heavily on the CDC (<http://www.pandemicflu.gov/>), provides links to state and local Web sites, and would be augmented during a crisis. Similarly, the Government of Canada maintains an interagency Web site ([http://www.influenza.gc.ca/index\\_e.html](http://www.influenza.gc.ca/index_e.html)) that relies heavily on Health Canada and the Public Health Agency of Canada.

Whatever the ultimate shape of the response to pandemic influenza, it will have the following elements, which should be part of any pandemic plan for employers:

- Social isolation. Efforts will be made to reduce the frequency of face-to-face contact among employees in order to slow transmission. Distance learning, telecommuting, and work at home will be encouraged.
- Risk and crisis communication. Efforts to inform and educate the public will be essential. (See Chapter 7.)
- Prioritization. Nonessential activities requiring human contact may be curtailed.
- Surveillance. The outbreak will be tracked by local public health agencies and internationally by WHO and CDC.
- Home care and triage. Persons showing symptoms of influenza will be encouraged to stay at home if their symptoms are low-

grade, may be cared for at designated intermediate care centers (which will be set up in schools and churches) if they require supportive care that they cannot get at home, and only admitted to hospital when severely ill. As in the case of SARS, H<sub>5</sub>N<sub>1</sub> influenza cases may initially be concentrated in a small number of hospitals in order to limit exposure if the outbreak appears small.

- Vaccine development. The vaccine will be developed as soon as the antigenic determinants are known but will take months to produce in quantity.
- Immunization. Immunization rests on three strategies. The primary strategy is that as soon as a specific vaccine is available, it will be administered as widely as possible, with priority given to certain groups (such as health care workers) for the earliest batches. The secondary strategy will be a polyvalent influenza vaccine that is designed on a probabilistic basis to cross-react with the strains of H<sub>5</sub>N<sub>1</sub> most likely to be present in the pandemic. The supportive strategy is to encourage near-universal immunization against seasonal influenza, so that during a pandemic the natural background of seasonal influenza can be minimized, removing a burden to the health care system and minimizing confusion. It should be possible to arrange mass immunizations for employees. The staff of the occupational health service may also be asked by local public health authorities to participate in community immunization campaigns if needed and if liability can be managed.
- Antiviral medication. There may be a limited role for oseltamivir (Tamiflu®) and possibly zanamivir (Relenza®) in a pandemic outbreak, especially for the seriously ill. Zanamivir must be taken by inhaler because it is not bioavailable orally, which limits its market acceptability. Supplies of both drugs are finite, and treatment inhibits the development of immunity, so that patients remain at risk for recurrent viral infection. Unfortunately, some avian strains are already resistant to oseltamivir.

Whether or not there will be a pandemic remains to be seen, although it seems likely. The value of pandemic preparedness however, is not limited to one threat. The principles are broadly applicable to other highly contagious emerging infections (such as SARS) and to bioterrorist threats (such as smallpox). The general lessons (the importance of surveillance, information, and interruption of the chain of transmission) may have application in other types of distributed emergencies.

The response to pandemic influenza has been criticized as disproportionate to the real threat. Given the disruption and mortality of pandemics in the past, however, failure to plan seems foolhardy. Shortly before 2000, in what was called the “Y2K” scare, many responsible professionals believed that the information systems of the world were at risk of paralysis because 2000 would be confused with 1900 in old programming formats. Although nothing happened, the desperate preparations for Y2K had the effect of replacing and upgrading infrastructure and making business IT systems, and the Internet, more resilient in advance of the emergencies of 2001 and more resistant to hackers. Similarly, pandemic preparedness will motivate and concentrate efforts in emergency management.

## **TERRORISM**

The threat of a terrorist attack on a private facility seemed remote until 2001 but since then has grown more plausible.

Terrorism has emerged as a threat largely as the result of converging issues of security, international relations, and public health. These include the following:

- Changes in the international order
- Changes in the ideology and practice of terrorism
- Changes in the capacity of the system to cope

Terrorism is a response to asymmetry of strength. Terrorists will attempt to use whatever tactics they can to make up the difference in force in order to get what they want. Terrorist groups have been changing tactics away from the traditional reliance on causing fear toward more “direct action.” This strategy seeks to deliberately incapacitate or seriously harm major institutions, such as financial system in New York in 1993 and 2001, rather than primarily to “send a message.” This is not to say that fear and terror do not still figure into the strategy, but terrorists are changing their strategies from fear and terror as primary objectives to one of using fear and terror as enhancement factors. Critical economic sectors and infrastructure are likely to be at higher risk than in the past, whether they are public or private. The potential for terrorism to become an occupational hazard has become significant.

Weapons of mass destruction have become a viable tool for well-organized international terrorist networks. The major reason to use them would be to do harm, to inflict injury, and to paralyze economic systems in ways that a conventional explosive, with its limited blast area, could not achieve. In addition weapons of mass destruction, particularly the ones that are most feared—chemical, certain biological agents, and radioactivity—increase the psychological impact. The public is more likely to stay intimidated if people perceive that they may be a victim of a mass casualty at any time rather than the unlikely victim who happens to be in the wrong place at the wrong time in a bombing.

Weapons of mass destruction are of five types:

- Chemical
- Biological
- Radiological
- Nuclear
- Explosive

Chemical agents have technical limitations that make them less effective for sophisticated terrorist operations. If the perpetrators have the expertise and access, they will undoubtedly use them, but they have limitations. The technical difficulties with using chemical agents largely reflect problems of dispersing them and rapid dilution in the atmosphere. Chemical agents are most effective in confined spaces and are unreliable even there. (The sarin gas attack on the Tokyo subway by the Japanese cult Aum Shinrikyo in 1995 only killed twelve people.) Once a chemical assault is over, it is largely over—chemical residues tend to degrade, adhere to surfaces, or dissipate quickly with time. The suddenness of the attack and the technical issues associated with dispersal (which requires the perpetrator to manipulate the device up to the last minute) also make it difficult for the perpetrator to escape.

Biological agents have a major advantage in cost, adaptability, flexibility, ease of concealment, and the impact on the population when they are dispersed. As a consequence they have emerged along with chemical weapons as one of the categories of choice for well-organized terrorist groups. Biological agents also have some potential disadvantages. There are sufficient technical problems in dispersing the agents to make bioterrorism difficult for most perpetrators. Clues may also be abundant in biological attacks because of the genetic material of the organism, which ultimately led to identification of the perpetrator of the anthrax assaults in the United States in 2001. Anthrax remains the most likely agent for a biological assault because of its favorable biological properties and stability of the spores.

Radiological agents contain radionuclides that during an explosion with a conventional charge contaminate an area with radioactivity. This is called a dirty bomb. The effects of a dirty bomb would actually be limited and mostly psychological. The immediate damage would be the result of the explosion, not the radioactive material, and would be limited to the blast zone. There would be no immediate health effects or acute symptoms because the radiation would be dispersed and so radiation poisoning would be unlikely. People in the

area and downwind would be at risk for radiation health effects over the long term, including elevated cancer risk, but the effect would not be seen for many years, and by then many people would have died of other causes.

Nuclear explosions, on the other hand, would be devastating and almost impossible to manage within the blast zone. Radiation fallout would be much more intense, and the blast effects would be far greater. There is no question that a nuclear threat is the most serious of the weapons of mass destruction.

Conventional explosives are often overlooked as weapons of mass destruction, but they can be serious threats to large buildings and public spaces. As a practical matter, when a blast occurs it is important to move away from windows and face the other way to avoid flying glass. It is also important to remember that perpetrators often place second bombs in order to catch the unwary, injure first responders (mostly police), and amplify the psychological effect.

A more detailed discussion of terrorist threats is beyond the scope of this book.

## RESOURCES

- Ciottone GR, Anderson PD, Auf Der Heide E, Darling RG, Jacoby I, Noji E, Suner S. *Disaster Medicine*. Philadelphia: Elsevier Mosby; 2006.  
Haddow GD, Bullock JA. *Introduction to Emergency Management*. Boston: Butterworth Heinemann; 2003: 37–47.

## NOTEWORTHY READINGS

- Brinkley D. *The Great Deluge: Hurricane Katrina, New Orleans, and the Mississippi Gulf Coast*. New York: William Morrow; 2006.  
Barry JM. *The Great Influenza*. New York, Penguin, 2004.  
Cahill K. *Emergency Relief Operations*. New York: Fordham University Press; 2003.  
Guidotti, T.L. (1986). Managing incidents involving hazardous substances. *Am J Prev Med.* (2): 14–154.  
Guidotti TL. Occupational medicine in time of crisis. *J Homeland Secur Emerg Manage*. 2005;2(3):art. 4. Available at <http://www.bepress.com/jhsem/vol2/iss3/4/>. Accessed September 30, 2005.

## The Praeger Handbook of Occupational and Environmental Medicine

- Guidotti TL. Managing incidents involving hazardous substances. *Am J Preventive Med.* 1986;2:14–154.
- Guidotti TL, Hoffman H. Terrorism and the civilian response. In: Upfal MJ, Krieger GR, Phillips SD, Guidotti TL, Weissman D. Terrorism: biological, chemical, and nuclear. *Clin Occup Environ Med.* 2003;2(2):169–180.
- Hogan DE, Burstein JL. *Disaster Medicine*. Philadelphia: Lippincott Williams & Wilkins; 2002: 108–109.
- Hudson TW, Roberts M. Corporate response to terrorism. In: Upfal MJ, Krieger GR, Phillips SD, Guidotti TL, Weissman D. Terrorism: biological, chemical, and nuclear. *Clin Occup Environ Med.* 2003;2(2):389–404.
- Keyes, Burstein JL, Schwartz Swienton, eds. *Medical Response to Terrorism: Preparedness and Clinical Practice*. Baltimore: Lippincott Williams & Wilkins; 2005.
- Landesman LY. *Public Health Management of Disasters: The Practice Guide*. Washington, DC: American Public Health Association; 2001: 145.
- Leas RD. Cleveland East Ohio Gas Company shows value of well-organized emergency medical program. *Ohio State Med J.* 1944;40:1172–1174.
- Levy BS, Sidel VW. *War and Public Health*. 2nd ed. New York: Oxford University Press; 2008.
- McLellan RK, Deitchman SD. Role of the occupational and environmental medicine physician. In: Upfal MJ, Krieger GR, Phillips SD, Guidotti TL, Weissman D. Terrorism: biological, chemical, and nuclear. *Clin Occup Environ Med.* 2003;2(2):181–190.
- Panamerican Health Organization. *Management of Dead Bodies in Disaster Situations*. Washington, DC: PHO; 2004.
- Vogt BM, Sorensen JH. How clean is safe? Improving the effectiveness of decontamination of structures and people following chemical and biological incidents. Oak Ridge, TN: Oak Ridge National Laboratory; 2002. ORNL/TM-2002/178

# **23 MEDICOLEGAL SERVICES**

The specialized health and medical knowledge in occupational and environmental medicine (OEM) are essential to the resolution of many health-related disputes in law and adjudication (such as workers' compensation) and provide essential input into public policy decisions, such as standards setting. OEM physicians often play an important role in evaluating scientific knowledge and defining the medical facts and interpretation of a particular case in law. This is a proper role for the qualified physician and a necessary part of OEM practice at an advanced level. Providing sound advice in legal disputes and administrative judgments is one of the most important services OEM physicians offer society.

This chapter is not intended to provide legal advice and was not written by a lawyer. Law and the rules of evidence, which govern what the court will accept or reject as valid to take into consideration, change over time and vary from jurisdiction to jurisdiction. It is essential that any OEM physician or any other expert be guided by the laws and rules of evidence in the jurisdiction where they are working or will be providing testimony. That means listening to the lawyer who is steering the case and not relying on nonlegal sources, including this chapter, as specific guidance.

In the United States and Canada, OEM physicians are involved in litigation as expert witnesses mostly in personal injury (tort) cases and occasionally in criminal cases (usually involving toxic exposures). Even more often, OEM physicians are involved as experts in the deliberations of adjudication systems derived from personal-injury law; such deliberations substitute the judgments of a panel, judge, or tribunal for those of a trial. They are called “alternative dispute resolution” systems. The most important and the oldest such system, certainly in North America, is the workers’ compensation system, which is at the center of OEM practice.

On a practical level, the legal system lacks the technical capacity to evaluate the validity of knowledge as evidence and therefore relies heavily on expert opinion. Much expert-witness testimony is disputed and controversial, and some experts are accused of crossing ethical lines and offering up “junk science.” However, the court depends on expert witnesses to help them interpret the facts of the case: it cannot function without them.

Expert knowledge—especially health and medical knowledge—is essential to the resolution of many disputes in law (such as tort litigation) and administrative adjudication (such as workers’ compensation). The medical or health expert provides essential input into public policy decisions. How to use this knowledge is not always clear. There are no broadly agreed-upon rules for the application of this knowledge, except those recognized by the law. In the United States, the culture of scientific investigation and the legal privileges given expert witnesses are reflected in legal precedents (cases that have already been decided by higher courts) and in the Federal Rules of Evidence. Although physicians are subordinate to the requirements of the legal system when they serve as expert witnesses, the law recognizes professional standards and the norms of medicine and, increasingly, epidemiology. The law allows testimony on professional standards to be treated as matters of fact rather than opinion, exempts the expert witness from the usual prohibition against “hearsay” evidence (when the witness did not directly and personally observe what happened), places great weight on the qualifications of an

expert, and is heavily influenced (definitively in states with the “Frye rule”) by the collective opinion of professionals. On the other hand, the law increasingly demands documentation and logic rather than opinion, and expects the expert to use the legal principle of the “weight of evidence” rather than the much more conservative scientific standard of certainty in interpreting the evidence.

## THE SOCIAL FUNCTIONS OF LITIGATION

Experts play a critical role in the resolution of some of society’s most taxing problems of equity and justice. The expert OEM physician is usually involved in complicated cases involving uncertainty, questions of causation, and arcane knowledge in medicine and toxicology. That is because the simple cases do not require such expertise. It is useful to review just why the OEM expert is doing this and what constructive social role is achieved.

For well over a century, people in the English-speaking world have thought that litigation is excessive and too expensive, as represented in Dickens’s novel *Bleak House*. The long, drawn-out lawsuit that is at the center of the novel takes place in a Victorian England that was rapidly becoming the first modern country: a mercantile, technology-driven, and commercially expansive contemporary superpower with new communities arising from the breakdown and formation of social classes, immigration, and internal population movement. Today’s times are similar, especially with respect to globalization and technological development; but the pace is much faster now, and disputes have to be resolved more quickly. The legal system is the ultimate venue where disputes are worked out, and if there is an expectation that a matter will land in court anyway, there is an incentive to litigate to get a definitive answer and skip efforts to negotiate.

The fundamental and ultimate set of rules for resolving disputes in a pluralistic society is the law. Litigation, the resort to law to resolve conflicts, is to be expected in a complicated, pluralistic, often-transient society. Societies that are homogeneous have other ways to settle their differences. Within traditional communities,

there are accepted informal means of dispute resolution, such as the intervention of village or tribal elders, religious figures, senior family members, “strong men” (government and business figures who have the power to impose a decision), private adjudication or arbitration procedures (increasingly popular in the United States), and public consensus (“the court of public opinion”). On the other hand, different communities do not necessarily agree about norms and acceptable behavior. In a pluralistic society where no one community dominates, the generally respected institutions for resolving disputes within one community do not apply to the others or to the majority. Some external means of dispute resolution is required. Mechanisms of dispute resolution that once existed within traditional communities also eventually break down in pluralistic societies, as cultures become more open, immigrants become assimilated, and society changes.

Even when people share the same culture and values, they may perceive the stakes as far too high to resolve disputes in a provisional fashion. The law is the final arbiter, and its decisions can be enforced. Why not go straight for the definitive resolution and save time?

Litigation also serves a deterrent function. Through anticipation and fear of an adverse judgment or a very expensive court battle, it deters the actor from actions that are likely to infringe on the rights or interests of others. It does so without being an absolute impediment, so some “gray areas” may be explored and risks can be taken within a narrow range. Compared to criminal law and to government regulation, neither of which are very adaptable or flexible—for good reason—tort liability emerges as the most flexible and potentially the most effective form of dispute resolution in law that our society has to offer, despite its high cost and inconvenience. If society is dissatisfied with it, at least part of that dissatisfaction can be blamed on the failure to support other ways of resolving problems.

Litigants contribute to a resolution of social problems related to responsibility and liability, contribute (occasionally) to setting precedents, and support the deterrent effect, all in the process of resolving their own cases. The litigants also bear the cost for this socially useful

function and, in some cases, the risk of catastrophic judgments. One problem with the social role of tort liability is therefore that the costs of providing a social function for all are borne by the litigants individually. The cost of this system is essentially underwritten by the losers in the actions and by the public, which maintains the judicial system. The costs may be large and unpredictable, and by the nature of litigation an action precipitated by one party may result in disastrous expenses for the other party, as well as collateral damage such as loss of reputation or business, even if they are not found culpable.

## TORT LITIGATION

The area of law that deals with personal injury—when one person or entity has caused harm to another—is called “tort law” in the American and Commonwealth legal systems. The injury may be a physical injury, an economic loss, a failure to fulfill a duty owed, or some other privation. This injury is called a “tort” in English. The French word from which “tort” derives is the same as the root for the word “torture” and the word “tortuous.” Many people think that this is highly appropriate, because tort law has often been difficult, painful, and convoluted, especially in the United States.

In recent years, class action suits have raised litigation to new levels of intensity and magnitude. Actions taken against asbestos manufacturers, the tobacco industry, and gun manufacturers have conferred a major social benefit, both by compensating the injured and by deterring further abuses. The huge judgments rendered and the legal fees paid have subsidized exploratory litigation by entrepreneurial law firms to open new grounds for tort litigation. The threat of a big lawsuit is a deterring factor against irresponsible or reckless behavior. At a time when the prevailing political sentiment was against regulation, and regulatory activity by OSHA and the FDA was restricted by Congressional intervention, this litigation may have served as the backstop to ensure that public health gains were not lost. Other cases have led to unfortunate results, such as the alleged fraud committed in high-profile class action lawsuits involving

silicosis in 2004 and 2005. Such cases not only abuse the legal system; they also compromise the chances of justice for legitimate cases and damage the credibility of OEM.

The application of medical knowledge in tort litigation has had successes and failures. Litigation over the legacy of asbestos exposure remains highly controversial: arguments over criteria for recognizing asbestos-related diseases are at the heart of the controversy, despite decades of high-quality research. Several efforts have been made in the United States Congress to set up an adjudication system, most recently in 2007. One particularly controversial issue has been the formulation of fair criteria for accepting claims. Litigation over silicone breast implants and the risk of autoimmune disease came to a halt after it was finally decided that the scientific evidence did not support claims of injury. Litigation over the safety of Bendectin (an antiemetic used in pregnancy) forced the drug off the market despite its proven efficacy and good safety record. The key case led to a major decision on the treatment of scientific evidence in federal courts. The common factor in these sets of cases has to do with opinions regarding causation, where these issues are particularly evident.

Much of tort law in the United States as it is applied to the risk of disease, rather than traumatic injury, arose from litigation before and during the 1920s. Plaintiff's counsel often attempted to relate cancer and other chronic health problems to physical trauma. A common example was the claim that cancer was the result of a recent blow or injury to a body part. The rationale invariably was one of *post hoc, ergo propter hoc*, meaning that because the outcome came after the incident, the incident must have been causal. As the science of the day became more sophisticated, courts soon threw out this argument as specious because there was no plausible mechanism relating the injury to cancer risk.

Occupational health- and injury-related tort litigation has a special burden. The workers' compensation system handles the vast majority of occupational cases—with varying degrees of success—for injury and for occupational disease. Because workers' compensation is an alternative dispute-resolution mechanism that constitutes

an “exclusive remedy,” with levels of appeal within the system but no appeal outside it, legal actions can only be undertaken on the grounds that the workers’ compensation process was unfairly applied (which is hard to prove in a quasi-judicial system that does not recognize precedent or follow the rules of evidence) or is unconstitutional to begin with (hard to argue given settled law and almost a century of acceptance). Needless to say, these legal actions are exceedingly rare. As a consequence, legal actions in occupational health usually involve “third-party liability,” in which the manufacturer or distributor is sued for failure to provide a safe product or to warn of hazards. Very rarely, cases of egregious neglect or willful disregard for workers’ safety will be prosecuted as criminal cases or even murder.

Environmental tort litigation is broader and highly diverse. A fuller elaboration is beyond the scope of this chapter.

In all matters related to procedure and timing, the expert should rely on legal counsel in the case to describe the steps in the process and the timetable. The sequence of events in American and Commonwealth legal cases normally involves filing a complaint, discovery, preparation of opinions, and depositions. Once the strength of evidence on both sides is apparent, there is a period of pretrial preparation, during which the possibility of settlement is usually raised. Very few cases actually go to trial. A decision to offer or accept a settlement may have nothing to do with the strength of the case, and the expert should never become emotionally invested in presenting his or her testimony at trial. It is not the expert’s case. Adjudication procedures are simpler and usually involve preparation of an opinion and possibly an administrative hearing.

### ***Frye and Daubert***

For many years, the judicial standard for expert testimony across the United States was the Frye rule (articulated in a 1923 decision in *Frye v. United States*, 93 F. 1013). The Frye rule was interpreted as admitting the conventional or consensus view of a particular field or

discipline, such as medicine, and discouraging admission of innovative or novel theories. This rule was based on wording from the opinion that stated, “the thing from which the deduction is made must be sufficiently established to have gained general recognition in the particular field in which it belongs.” Over time, the application of the Frye rule is believed by many to have drifted into admitting professional opinions that could be found to have some basis in some field, however narrow or self-defined. For others, however, the Frye rule, if rigorously applied, is a correction against individualistic or poorly supported opinion and a means of grounding testimony in the mainstream of the expert’s field of practice. The Frye rule remains the basis of evaluating or allowing expert testimony in many states, including California, but its application depends importantly on state law and precedent.

In the United States, a 1993 Supreme Court decision attempted to clarify the standard for applying scientific information to dispute resolution in federal courts and in many states that apply the Federal Rules of Evidence. The decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (507 U.S. 579) set forth a new and higher standard for federal courts in reviewing scientific evidence. (This was the decision in the Bendectin case.) The effect of this decision was that judges presiding over technically complicated cases were instructed to serve as “gatekeepers” and were required to evaluate the admissibility of evidence proffered by experts. This requires evaluating and monitoring scientific evidence, evidence that judges cannot be expected to have mastered. This decision was later reflected in the decisions of many state supreme courts, including that of Texas (which reached similar conclusions in the form of two mutually supportive opinions, usually called *Hawner-Robinson*).

In keeping with an earlier trend in some state appellate courts and in general trends in adjudication bodies, *Daubert* requires federal courts to examine the quality and methodology underlying testimony in arriving at their admissibility decisions and thus to apply the standards of science to scientific testimony. Its influence has been felt throughout the legal system, resulting in higher expectations for

rigor and persuasiveness in the opinions offered by expert witnesses, even when the local standard is the Frye rule or some variation of it. A consequence of the *Daubert* decision is that it is now much harder to demonstrate sufficient admissible evidence to support a “first case” when a hazard is new or an association has not previously been recognized.

The *Daubert* decision imposed a great burden on courts. Few judges and clerks are prepared to assess scientific data independently and few have staffs equipped to do this knowledgeably. Most lawyers will agree that law school was never designed to prepare them for technical issues in science. In practice, decisions on whether to admit expert testimony in a case may require a special evidentiary hearing to determine whether the expert has met the standards of *Daubert* or the applicable state rules of evidence.

Since *Daubert*, courts have required more documentation of the evidence and have set a higher standard. Peer review is now accepted as meeting this legal standard, and experts are often asked on the stand if the evidence they cite and the opinions, or theory of the case, they espouse have been peer-reviewed. Theories that are specific to a particular case have no opportunity to be peer-reviewed. The assessment of whether a scientific theory is worth hearing therefore takes place before the jury ever hears it or the judge allows it to become part of the record.

Other legal systems have taken other approaches. For example, in Canada the emphasis in the courts is not on excluding or evaluating scientific evidence when it is introduced. The emphasis is on weighing the testimony for scientific validity after it has been given. The jury, with instructions from the judge, decides whether what they have already heard was worth hearing.

## THE EXPERT WITNESS

Expert testimony is an old and venerable function of health professionals. In general, the law respects the opinion of physicians and other expert witnesses, but the role of the expert witness has changed

from providing an authoritative opinion alone. In past years, the informed judgment of health professionals, without reference to the evidence, carried greater weight than it does today. Junk science and the spectacle of “dueling experts” who seem to fight without firm grounds for their positions have provoked a backlash. Since the *Daubert* decision, courts have put much greater emphasis on defensible arguments based on empirical data and less emphasis on expert judgment. The ability to base testimony on evidence, and to fit the evidence together in a way that appears objective, is far more important in today’s courtroom. A personal, subjective, or idiosyncratic interpretation of the facts contributes little and may even undermine one’s professional credibility.

The role of the expert witness is to establish and interpret issues of fact. There are five essential responsibilities for the expert witness:

- To establish pertinent facts
- To place them in a context
- To explain the theory of the case in terms that are comprehensible to the lay court
- To render quantitative assessments in qualitative terms
- To derive an assessment of the degree of injury in terms that can be dealt with by the court

The role of the expert witness in civil cases is normally to provide testimony and informed opinion to establish or define issues of fact in a tort. The medical expert witness is also usually called upon in these settings to testify as to causation and damages related to health and physical impairment. Opinion is not enough, but it is allowed. This is one of the few situations in legal proceedings when witnesses are allowed to testify on the basis of personal opinion rather than knowledge. An adjudication proceeding, such as an appeal of a workers’ compensation claim, is similar.

The first responsibility is to present pertinent facts in a meaningful context. It is now more important than ever to show that these facts

are pertinent to the case and reflect the best scientific information available. It is also critical to integrate the facts of the discipline (such as cancer rates for various occupations and exposures) with the facts of the case so that the testimony is not an abstract discussion, but directly pertinent to the individual circumstances of the case. This also means careful attention when the profile of a particular plaintiff or claimant does not fit the average of characteristics or the overall profile of a population in an epidemiological study—for example, if a plaintiff developed cancer at an unusually young age. It is not enough to testify that there is an association between exposures characteristic of an occupation and the risk of cancer, for example. It is now important to specify that for workers of a certain age and background risk profile there is persuasive evidence that the risk is elevated to an extent that for this group it is more likely than not the cause of the site-specific outcome in any particular case. This is the same logic as a “rebuttable presumption,” which will be described later in this chapter.

Explaining the context is the second responsibility. The meaningful context for these facts is just as important as the facts themselves. The facts must be related by scientific interpretation in the expert witness’s treatment, just as the elements of the case are related by legal principles in the lawyer’s analysis. The expert witness has always been expected to express a sound opinion in a comprehensible fashion. However, the expert is now expected to provide solid grounds and a coherent chain of logic for the opinion expressed, and to place it in a context that assists the adjudicator in arriving at an informed decision. The expert is expected to reflect either a professional consensus or a well-supported minority opinion with evidence of backing in the scientific community.

This interpretation must be explicit and convincing. The expert needs to explain why an exposure probably caused the outcome in the particular case, establishing what is called “causation,” described in greater detail below. This may mean presenting the mechanisms of carcinogenesis in more or less detail to make the link clearly. It is necessary to establish the clear chain of logic and

to base the interpretation of the case on the best available facts and estimates, but unless these scientific considerations eventually lead to a qualitative assessment of “more likely than not,” they are going to be more confusing than helpful. Thus, the expert witness still must render an informed, qualitative opinion, but it must be preceded by and grounded upon a defensible scientific argument. The expert has a responsibility to provide a clear rationale behind the opinion and to articulate it in a manner that is useful to the adjudicating body.

The third responsibility is to communicate the theory of the case in terms that the judge or jury has no trouble grasping. The cliché is that judges, like most lawyers, went into law because they did not do well in science at school, and sometimes this is true. Finding a theory that makes sense is not enough. The theory of the case must also be communicated clearly and simply. In court, even a reasonable theory developed to fit the particulars of a highly unusual case may appear idiosyncratic, even bizarre, when distorted by the “other side.” “Framing” is a term of art for reviewing the facts of the case, developing a theory of what happened, and explaining what happened as a coherent narrative. A robust theory should stand up to more than one interpretation of events and does not depend on too many assumptions; it should never depend on just a single fact. This is not easy, but if a theory of a case is likely to be true, it should be able to explain all of the essential facts of the case and predict other facts as well.

A common mistake on the part of plaintiffs’ experts is to focus narrowly on proving the case and then to go no further when a reasonable theory has been formulated. Experienced experts focus on disproving the case and finding what is wrong with the theory, whether they are advising the plaintiff or defense. For the plaintiff expert, it is essential to put the theory to the test before the opposition does so in court. For the defense, the task is much easier. The defense expert only needs to poke holes in the plaintiff’s theory and raise the possibility of other explanations, not prove them.

An apparently rational but complicated theory of a case may easily require so many contingent steps that the final odds that the theory resembles what happened are much less than even; in other words, the more elaborate a theory, the more likely that part or all of it is wrong. Explaining a theory that complicated to a jury is also not easy. The simpler the theory and the fewer the steps, the more likely it is that the entire sequence of events is true, and the easier it is to explain. (This principle is called “Occam’s razor.”) So the essence of framing a theory of a case is to identify the key facts, put them together simply, challenge the theory to explain anomalous or unexpected facts (before the opposition does this in court), and gather the supporting evidence to document the theory.

A critical part of clear communication, and the fourth responsibility, is to render quantitative information in qualitative terms. Numbers should be kept to a minimum in expert testimony, especially in a jury trial, but whenever possible they should be described with examples. A probability of 50 percent is an abstraction; the balance of probabilities is a metaphor; a coin toss is a vivid example everyone can understand. It is not enough to speak of a probability of  $10^{-6}$ . The concept of one in a million has to be made graphically, such as by referring to one person in the whole population of Denver or Calgary or some other comparison that will mean something to a person without technical training. A part per billion has to be compared to ten drops of water in a swimming pool, not expressed as  $10^{-9}$ .

In some tort cases and in many workers’ compensation adjudication proceedings, the expert also has the role of evaluating damages, describing future health risks, and rating disability. Evaluating the degree of injury in terms that the court or tribunal can understand may be of greater or lesser importance depending on the case and the qualifications of the expert witness. For medical experts, this may involve an assessment of permanent impairment or long-term disability and future risk of developing a complication. On the defense side, it may involve a rebuttal. This aspect of the expert witness’s role is highly individualized and reflects the circumstances

of the particular case and the expert witness's specific role as determined by counsel.

The adversarial structure of the American and Commonwealth legal systems encourages extreme interpretations. This has given space for junk science and bogus experts. One might ask why the court does not use its own experts. Something like this has been attempted by judges who have set up expert panels to advise them in class action lawsuits, but such examples are very rare. In the rare instances when a judge has had access to a consultant capable of rendering an independent assessment, there have been concerns that the in-house expert could unduly affect the decision by manipulating the assessment and by inadvertently supplanting the role of the judge in deciding on the admissibility of scientific testimony.

The foundation of a trial in the American or Commonwealth system of law is advocacy on behalf of two opposing sides that have a duty to present their own best case. It would be inimical to the operation of the legal system if, for example, plaintiff and defense experts, or claimants and adjudicators, let their experts meet in conference to decide among themselves which scientific interpretation is correct.

## **CAUSATION**

Most cases in which an OEM physician is likely to be engaged involve causation. Properly speaking, causation is the causal event, and causality is the relationship between cause and effect: one is specific and the other is general. However, they are often used interchangeably, with "causality" being preferred in the United Kingdom and "causation" in North America. The systematic process leading to the determination, to a reasonable approximation, of the most likely cause is called "causation analysis." For the OEM physician, causation analysis requires a familiarity with the methodologies of epidemiology, toxicology, clinical medicine, and,

increasingly, risk science. (Causation and causality in epidemiology are discussed in Chapter 3.)

“Causation” refers, in this context, to the risk factors or exposures that initiate the process leading to the health outcome. The concept is akin to that of etiology in clinical medicine, but without the implication that there must be only a single cause. The concept of causation in epidemiology assumes that the risk factors bear a “causal” relationship in that they either establish a necessary condition that allows the outcome, or set into motion a mechanism that results in the outcome. There is no assumption, as there is in common language and in the work of some epistemologists, that a cause must be “sufficient” in itself to produce an effect.

A cause, in the sense that is used in most toxic tort cases (especially when cancer risk is involved), is a factor that contributes to the likelihood that an outcome will occur. This is a stochastic, or probabilistic, definition, not strictly a mechanistic definition. There is a certain probability, or odds, that a step will occur, but no certainty. In daily life, one speaks of “cause and effect” relationships as if there is one cause for every effect and as if an effect necessarily follows the presence of a cause. This is too rigid to be useful in epidemiology and especially in carcinogenesis, where the mechanisms are complicated and influenced by numerous external and internal factors. It is not even useful, in this context, to speak of a cause as being either necessary or sufficient, because causes may be interchangeable in the mechanism or may interact.

For example, exposure to either cigarette smoking or asbestos individually is known to result in lung cancer in a roughly predictable probability. Exposure to both increases the risk beyond that of the summed probabilities of either alone, suggesting a substantial interaction. However, most workers who have been exposed to either or both do not develop lung cancer, although they might if they live long enough and are free of other risks to their life. A few unlucky people who neither smoke nor become exposed to environmental carcinogens such as asbestos develop lung cancer

regardless, although this is uncommon. Neither asbestos exposure nor cigarette smoking is necessary, sufficient, or predictable in individual cases as a cause of lung cancer, but the association is clear and these factors are truly “causes.”

The medical expert needs to review the case with an eye toward many factors in causation. These include:

- Determination of the proximate cause (what was immediately responsible)
- Determination of the underlying causes (what conditions created the situation)
- Liability for the causal situation (who was responsible for allowing it to happen)
- Contributing behavioral factors
- Apportionment among the various causes that are possible
- Preexisting conditions (such as atopy) and personal factors (such as host defense or susceptibility states) that confer susceptibility (a biological condition that makes effects more likely)
- Occupational, environmental justice, and behavioral factors that confer vulnerability (a social condition that makes a person more likely to be exposed or disadvantaged in protecting him- or herself)

Causation is similar to the concept of etiology in clinical medicine but without the implication that there can only be one cause. In clinical medicine, etiology is usually not as important as diagnosis, because in clinical practice, regardless of what caused the condition, the task at hand is to treat the patient. In law and policy, however, the assessment of causation is critical because it establishes liability and responsibility. Causation is even more central than diagnosis in initiating compensation and stimulating prevention. It is not at all unusual for a physician to be able to make a diagnosis without knowing the cause. It is also not always necessary to know the exact diagnosis to

assess causation, as long as the pathological process is known. For example, one can know that a worker has nonallergic airways reactivity following an exposure clearly associated with work without determining for certain whether it is occupational asthma associated with irritant exposure, reactive airways dysfunction syndrome, or aggravation of preexisting but subclinical airways reactivity in allergic rhinitis.

The concept of causation has different implications in different disciplines. In epidemiology, the concept of causation assumes that the risk factors bear a “causal” relationship to each other in that they either establish a necessary condition or set into motion a process that results in the outcome. This does not necessarily mean, as in common language, that a cause must be “sufficient” in itself to produce an effect. In toxicology, the concept of causation tends to be more mechanistic, and assumes that there is a chain of biological causation from first exposure to outcome. However, toxicologists are also accustomed to thinking in terms of complex interactions. In OEM, causation is the determination of the most-probable cause of the worker’s condition or disability. In workers’ compensation, to establish causation means identifying the factor that created the condition and also demonstrating that it arose out of the work setting or conditions.

In law, there are two elements of causation: cause in fact and proximate cause. Cause in fact is the necessary condition, similar to the notion of “underlying cause” on a death certificate. It can be identified by a simple test: would the adverse outcome have been avoided “but for” the presence of the cause? A cause in fact can be necessary but not sufficient. The proximate cause is the particular event or factor that initiated the chain of events leading to injury or damage; it is comparable to the “immediate cause” on a death certificate. In law, especially in adjudication proceedings, a proximate cause may also be a “substantial factor,” which contributes to the outcome even if it is not wholly responsible. Thus, in a heavy cigarette smoker who was also exposed to large amounts of asbestos and developed asbestosis, and who subsequently developed lung cancer, the smoking may be a

cause in fact, but the asbestos exposure could be a proximate cause and the basis for a judgment. This is because not all or even a majority of cigarette smokers get lung cancer (although they all might if they survived long enough), but the effect of asbestos exposure on a cigarette smoker more than doubles the risk of lung cancer: so most cases would not have arisen but for the asbestos exposure.

A good medical evaluation alone is not enough for causation analysis. Causation analysis must be undertaken methodically on two levels. The individual case must be thoroughly documented and evaluated. Then, the evidence in the individual case must be linked with the broader body of knowledge in science regarding risk, usually derived from epidemiology, taking into account how the individual may differ from or resemble the populations studied.

One systematic approach to causation analysis is the following:

- Confirm the diagnosis and the nature of the injury. Occasionally, a report of the diagnosis will be incorrect.
- Determine the circumstances of exposure. Confirm that there was a plausible route of exposure of sufficient magnitude to place the person at risk.
- Determine other possible exposures that may be causal or significant in the case and their plausibility and likely magnitude. For example, cigarette smoking is by far the most common alternative risk factor in occupational lung-disease cases, so the smoking history is often critical.
- Determine whether the injury is a recognized effect of the exposure. This is relatively easy to do for carcinogens because the International Agency for Research on Cancer is an authoritative body that provides assessment of the evidence for carcinogenicity of chemicals and industrial processes. It is more difficult when the scientific literature is incomplete and may be impossible in a “first case.”
- Determine whether the particulars of the individual case are consistent with findings in the scientific literature. This often

requires a working knowledge of toxicology because outcomes may differ in human and animal studies.

- Develop a reasonable hypothesis for how such an event might have happened—the simpler, the better. Then test this hypothesis against the facts of the case. Does the theory predict a different fact that can be checked?

Causation may be simple if the disorder is typically associated with a single external cause, such as mesothelioma and asbestos exposure. The challenge in such cases is to demonstrate the opportunity for exposure to an agent that is a known cause, which requires a thorough history. Medical records are usually poor sources of the exposure history because the first person to take the history is rarely trained to obtain a thorough occupational history (see Chapter 15), is not primarily interested in exposure assessment, and is relying on second-hand information from an anxious patient. Subsequent references to the exposure in the medical history often merely copy the first and occasionally distort it, perpetuating errors in the process.

Conditions for which there is a strong or suspected association with occupation on a group basis may be hard to prove on an individual basis. Some workers' compensation systems have schedules of recognized diseases, for which claims qualify more-or-less automatically if there is a history of exposure or if the worker is in an occupation with a demonstrable and accepted risk. However, the association may be disputed in the individual case if there is evidence for another cause, evidence that exposure was not sufficient, or evidence for some other mitigating circumstance. This is called a rebuttable presumption and is an important policy alternative in workers' compensation and legislation. Presumption is discussed later in this chapter.

Identifying a particular cause for the outcome is often difficult. This is illustrated by common occupational lung diseases. The occupational and environmental history certainly narrows down the range of possible exposures and may rule out all but one or a few possibilities, particularly in the case of the pneumoconioses, which

never occur without exposure to the causal dust. However, cases of asthma and hypersensitivity pneumonitis are rarely so easy, because the disorders may mimic common nonoccupational and the number of possible exposures triggering the immunologic responses may be large. Toxic inhalation cases are always associated with a discrete event; there may be uncertainty over which particular chemical exposure caused the injury, but the association with the event in question is rarely in dispute. Obstructive airways disease and cancer are exceedingly difficult and often impossible to associate with a particular exposure in the presence of other risk factors, such as cigarette smoking.

In many cases, the OEM physician has available a variety of records to develop an inventory of exposures and possible causes:

- A thorough occupational history
- Employer personnel records
- Industrial hygiene survey records (when available)
- Union records
- Pay stubs, tax records, and income statements that identify particular employers
- Medical records
- Testimony of co-workers
- Past workers' compensation claims

At other times, the medical record and most recent occupation is all that is available. When the information is insufficient to come to a firm conclusion, the expert should say so and then use the best judgment and most reasonable assumptions possible, spelling out in detail at every step what assumptions have been made and the basis for them.

The evaluation of occupational and environmental disease may be assisted by more-advanced diagnostic modalities. Technology has advanced the precision of medical diagnosis, but it does not necessarily

help to the same degree in causation assessment and etiological determination. The use of advanced diagnostic tests without guidelines may even bias the system. A claimant who receives sensitive but not specific tests, such as high-resolution computed tomography for fibrotic lung disease, might be more likely to have his or her claims accepted than a claimant whose physician did not use that technology, simply because it looks more convincing to a judge. Likewise, x-rays of the spine that show erosion of the vertebral body and spurring are often accepted as evidence of a structural abnormality for low-back pain alleged to have arisen out of work, yet such lesions are common in people without low-back pain and are not by themselves evidence of an occupational cause. Therefore, diagnostic tests are not always helpful and may sometimes confuse the picture.

The evaluation of occupational disease may appear very conservative to practitioners who are accustomed to applying the latest medical breakthroughs, but there are good reasons for this. Workers' compensation evaluations tend to rely on studies with known and familiar technology. Traditional tests, such as the chest film, have been exhaustively studied in occupational medicine and are well understood; for example, progression of asbestosis on a chest film has a relationship to lung function and cancer risk, but this has not been established for high-resolution computed tomography (HRCT). Insurance carriers may not authorize newer or highly sensitive tests that are not easily interpreted. The history of OEM is littered with nonspecific, unhelpful, or misinterpreted studies: trace element analysis (see Chapter 2), PET (positron emission tomography) scans used for "toxic encephalopathy" (see Chapter 1), serum mold antigens and antibodies, visual contrast sensitivity, and a raft of tests associated with unproven medical theories, such as multiple chemical sensitivity and its progeny. There have also been egregious cases of questionable tests being applied in fitness-for-duty evaluations and periodic health surveillance, where incorrect and invalid results had the potential of costing a worker his or her job or career. This is why the issue of validated testing is so sensitive in health monitoring (see Chapter 5) and OEM ethics (see Chapter 25).

Another peculiarity of causation analysis compared to conventional medical assessment, diagnosis, and management is that treatment records are generally useless. Unless the clinical management of a condition is part of the question being put to the medical expert, as in medical malpractice, nothing that occurred after the disorder appears is of much interest in causation analysis. Treatment records should still be scanned by the expert, however, because every once in a while the later notes reveal a useful fact. Likewise, nursing notes are usually not very informative, except in reporting the patient's state of mind and anxiety, and sometimes introduce or perpetuate errors, especially in matters of diagnostic detail and smoking history. On the other hand, pulmonary function tests incorporate a short smoking history, which may be more accurate than what is found in the history of present illness.

A finding of causation is a reasonable conclusion when a case conforms to risk factors that are described in the scientific literature, and there is a clear consensus that the exposure is associated with the disease outcome. These are the easier cases. The harder cases are those in which the literature is not clear. The first step is to review the literature to determine whether the evidence suggests a causal association. The literature may be clinical, toxicological, or epidemiological. In general, the clinical literature tells the OEM physician that something happened, the toxicological literature tells the physician that something could happen, and the epidemiological literature tells physician that something does happen. None of them can tell the physician what happened in the specific case; that is a matter of inference.

As a practical matter, empirical evidence that can be demonstrated to apply to a given case trumps a mechanistic argument that is based on toxicological principles, which in turn trumps personal opinion based on clinical experience. Toxicological studies may demonstrate that an association between exposure and an outcome in the case exists in animals and may demonstrate a plausible mechanism. The methods, relevance to humans, comparative exposure levels, species, and strain become matters of interpretation.

It is important to know whether the profile of the injured party/worker corresponds to the population the study has investigated. It may matter, for example, that the plaintiff/claimant is younger, has worked longer, does or does not satisfy the expected latency period, or belongs to a particular subgroup in the analysis. An important principle in workers' compensation is to "take the worker(s) as they come," meaning that once they are hired, personal characteristics do not matter. For example, unless there is a policy against hiring people who smoke, an employer cannot argue that an exposure that makes only smokers sick is not work related. Some claimants, by their family history or some acquired condition, may be more susceptible than others to injury. Such persons are known as "thin skull," "eggshell," or "eggshell skull" plaintiffs after an early legal case in which a person sustained a skull fracture after a minor blow. The fact that the victim was more susceptible to a skull fracture than most people did not absolve the person responsible for hitting him on the head from liability and paying damages.

"Aggravational" causation occurs when an exposure makes an existing condition worse or brings out symptoms in a previously silent condition. Exposure to an irritant gas, including passive cigarette smoke, may aggravate reactive airways disorders in an individual with atopy, for example, even if he or she has had no previous asthma. This is not the same as de novo asthma but may be accepted as permanent impairment in some systems if it persists and interferes with the worker's ability to perform his or her usual occupation. Some systems use the term "exacerbation" to refer to a longer-term, often permanent, worsening of symptoms.

"Substantial contribution" is a feature of some legal systems. This doctrine suggests that the condition should be considered to have arisen out of the workplace if the exposure was sufficient to contribute more than trivial risk and would have added enough to the causation of the disorder to tip the balance, all other things being equal. This criterion is particularly common in situations in which strict causation is difficult or impossible to prove or in which the job was particularly hazardous or involved national security and the

assumption of personal risk. The Black Lung Benefits Program in the United States is such a system, administered under the Department of Labor (with some supplemental benefits from Medicare).

Some sense of apportionment by cause may be needed to explain the history and presentation of the disorder and to sort out the extent of contribution of work-related factors when the relationship is not simple. In a given case, there may be more than one plausible agent or risk factor responsible for the condition, such as cigarette smoking, exposure to asbestos, and exposure to other chemical carcinogens. Separation of the nonfactors, and attribution of the factors to the correct employer, is an ideal to be sought, but it is rarely feasible in practice. Demonstration of substantial contribution by workplace-associated exposures is usually sufficient to declare the disorder and all associated impairment to be work related. Since 2004, apportionment of cause has been required under workers' compensation law in California. It may be tempting to apportion based on attributable risk, but this is an error. Attribution is an estimate of the contribution of a given cause to disease in a population, based on its role in a proportion of cases; apportionment is an estimate of the contribution in an individual case, where the cause may be all or partial, or may not be a cause at all.

## **CERTAINTY**

Society uses many different criteria for evaluating risk, accepting what is likely to be true, and deciding to act. OEM physicians and others who are well prepared and formally trained know how to evaluate evidence in science and have internalized the "95 percent certainty" principle for statistical significance or a change in baseline condition. However, civil law (and most systems of adjudication) has a different standard for disputes between parties: the balance of probabilities, or "weight of evidence," which translates to greater than 50 percent certainty. When the medical or health expert ventures into the courtroom, therefore, it is like playing a (serious) game with very different rules. By comparison, the standard of certainty for criminal

cases, when the accused will be deprived of rights, is “beyond reasonable doubt,” a standard even more stringent than 95 percent certainty.

The usual criteria applied in civil litigation and for accepting an association as causal is that the disease in question is “more likely than not” associated with the exposure or act of the defendant. The same standard of “50 percent + 1,” or “balance of probabilities,” is also the legal criterion for judgment in a civil case, and in a civil case the burden of proof is on the plaintiff. In workers’ compensation cases, the benefit of the doubt is given to the claimant. One need only be “50 percent sure” (the exact balance of probabilities, when there is equal weight on both sides of the question) to make a judgment.

Statements of certainty are just that, not statements about relative contribution or apportionment. The expert helps the court to the extent possible to recognize a balance of probabilities where it exists on the basis of evidence. The court is looking for pivotal points, thresholds, or “bright lines” (a term of art meaning a clear distinction) on which to base a decision.

Various jurisdictions use different terminology to express the idea of certainty or a balance of probabilities. All terms are intended to express the idea that one is persuaded, by the weight of evidence, that the evidence for causation by the factor in question is at least as good or marginally better than the evidence for causation by any other factor and by all other factors in combination. These terms should be considered a different vocabulary of expression that is specific to the medicolegal or adjudication setting. For example, many jurisdictions expect the expert to use the phrase “to a reasonable degree of medical certainty” to express confidence in the opinion, as well as recognition that the science and art of medicine have limitations. If so, whether the expert likes it or not, the phrase should be used, because it is actually a signal that the expert understands what he or she is being asked within the context of that system. On the other hand, some terms are universal. For example, where the “50 percent sure” criterion applies, as it usually does, “probable” means yes and “possible” means no, because “possible” means less than 50 percent certainty to most people. No distinction is made between being very

sure and being certain; all that is important is that a conclusion based on fact is definite and more likely than not. To be unsure or uncertain is not to support the conclusion. It thus adds little to an expert witness's testimony to go into detail on the witness stand about minor reservations or to express an unlikely alternative opinion as a possibility. When a scientific expert is asked for an expression of certainty, there is a natural tendency to couch the statement in the usual scientific terms, with many qualifiers. This is a mistake. From the legal standpoint it is of no interest, but expressing such reservations may have a psychological effect on the jury or adjudication panel that may confuse the case.

Most adjudication systems give a nominal benefit of the doubt to the claimant in cases where the odds are evenly matched, but these systems zealously guard their right to confer this benefit of the doubt. The expert witness should not assume the prerogative of making this judgment. The role of the expert witness is to outline the framework for a judgment of probable or likely causation clearly and as rigorously as the discipline allows, and then to let the system decide.

Some medical expert witnesses stick to the familiar rules of science and are therefore, by definition, too conservative in their opinions. Others may feel liberated by the looser standard of civil litigation and free to make up theories and opinions that are extrapolated far beyond solid evidence. An example of this is suspect testimony in the wave of litigation over "toxic mold" in the United States today. Litigation has been a spawning ground for so-called "junk science," which has threatened the credibility of experts in general and has probably discouraged many knowledgeable investigators and practitioners from sharing their knowledge when it has been needed.

There is nothing unethical about holding one opinion with respect to the legal interpretation of a set of findings and another with respect to the scientific interpretation. One may legitimately consider a matter to be very likely but not scientifically proven (such as asbestos as a cause of colon cancer). Often, the scientific evidence for an association is strong but not conclusive. In such cases, it is

entirely reasonable and responsible for an expert witness to maintain on the witness stand that there is an association on the basis of an interpretation of “the weight of evidence,” but to maintain in a scientific forum that the association is not proven, because it has not been proven beyond scientific doubt. The reverse is also true: It is legitimate for an expert witness to believe, as a matter of opinion, that there is not an association but that the evidence against it is not conclusive. (This position, which might apply, for example, to cancer induced by electromagnetic fields, is very common because the doctrine of falsification makes it clear that a negative can never be conclusively proven.) The law is interested primarily in the weight of what evidence exists, and secondarily on how strong the body of evidence is in its entirety. A corollary to this is that the testimony of an ethical expert witness on the stand may not coincide with his or her professional opinion in a scientific setting. (Some sources cite a 1984 decision of the United States Court of Appeals for the District of Columbia—*Ferebee v. Chevron Chemical Co.*, 736 F.2d 1529—in support of this principle; however, this is incorrect. *Ferebee* addressed the sufficiency of evidence supporting an opinion rendered in court and the right of an expert to hold an opinion based on interpretation of the evidence. It did not address consistency or divergence of the same expert’s opinions or conclusions based on the same evidence outside of court and in the realm of science. That is not a matter for courts to decide, anyway.)

Testimony based on the latest findings and just-published studies gives the appearance of being up-to-date and irrefutable. Likewise, a common ploy in cross-examination is to cite a new or obscure study that takes the expert by surprise. (This is called the “gotcha!” strategy.) However, both tactics are actually weak because the latest findings have not had time to be evaluated and integrated into the body of scientific interpretation, which is what really matters, and the most obscure studies never will. Their probative value is therefore limited until they can be examined and, in most cases, confirmed. On the other hand, it is ethical to testify that there is an association on the basis of preliminary evidence while acknowledging that systematic

efforts to disprove the hypothesis have not yet been made. It all depends on how the information is used and how the argument is framed. In general, however, an argument that pivots on a very new or very obscure finding is weak. Paradoxically, these findings carry most weight when they are consistent with the body of science, not when they contradict it.

The courts only expect an interpretation based on the best evidence available, not prescience or omniscience. What counts most is the weight of what evidence exists, not how strong the body of evidence is in its entirety. The court cannot wait to make a decision until all the evidence one could want is available.

When an expert concludes that the balance of probabilities, or the weight of evidence, favors the conclusion that a particular hazard caused an injury or disease, the statement only refers to the relationship between the hazard (cause) and the effect. Often the factor in question is the most likely by a plurality, not a majority share. There may be other possibilities and the cause in question may be the most likely among them. That test is whether the likelihood that the outcome was caused by the hazard is the most likely among several possibilities (usually by a wide margin), but not necessarily at a level of 50 percent probability. The expert witness must then explain that the weight of evidence favors the most likely cause against all the other possible causes individually. If the hazard in question is at least equally probable to have been the cause as the combined probability of all other possible causes, then it is already at the balance of probabilities.

Statements of certainty can also be applied to incomplete causes. In this application, they describe how sure one is with respect to whether some factor was a cause at any level. Some adjudication systems find for the claimant if the hazard or action played a role or substantially contributed to the outcome. For example, the managed federal tort systems for railroad workers do not require a conclusion that the hazard caused an illness by itself; it is sufficient if it made a substantial contribution. However, one must still decide whether it is more likely than not that the factor in question contributed anything substantial to the causation of the outcome.

These statements of certainty can be read in one way as descriptions of whether a particular factor was the cause of a particular disorder. This is true only when the assumption is made that there is only one sole and sufficient cause of the outcome. There may be several possible causes, but the argument proceeds on the basis of which cause is more likely to be responsible. This is indeed a common situation, such as in cases disputing whether exposure to a carcinogen was sufficient to cause a cancer or whether there were other causes that could have caused it. Here it is enough to determine that the evidence favors the cause in question being sufficient and more likely than any and all of the other possibilities.

A related issue arises when exposure to a cause is necessary but not sufficient. For example, if a particular exposure raised the risk of a heart attack, but only in persons with high blood pressure, it should be enough to demonstrate that the available evidence shows it to be more likely than not that “but for” the exposure, the person would not have had a heart attack, regardless of the cause of his or her underlying hypertension.

The relative emphasis to be placed on each of these partial roles depends on the case and the weight of available evidence. The more the facts of the case speak for themselves, the less important it is to interpret them and the less work the expert witness must do to establish their validity. As the weight of evidence comes closer to certainty, the burden of interpreting the evidence correspondingly becomes lighter. The expert’s role then becomes to explain and present the evidence clearly and accurately.

Judgment remains important in the testimony of the expert witness. Indeed, the expert witness is excluded from the hearsay rule in presenting evidence precisely because he or she is deemed qualified to present an informed opinion based on the collective wisdom of the discipline. However, where judgment is needed, the reasons for making an informed judgment one way or the other should be spelled out. The areas of gray, the reasons for choosing one estimate over another, and the means by which one deals with uncertainty should be made explicit.

Gold (1986) defined three levels of certainty in litigation:

- Burden of proof
- Preponderance of evidence
- Standard of persuasion or belief

The three levels are frequently confused. In particular, the preponderance of evidence and the standard of persuasion are often “collapsed” (in the terminology of Gold) into a single statement of how convincing the evidence may be. It is more useful to think about them separately. Table 23.1 presents the three levels and key descriptors that anchor statements of certainty for each.

Causation—did the exposure cause the damage or not?—is a dichotomy, a true or false proposition. In order to decide, the pertinent facts must be presented, and this burden falls to the plaintiff or claimant. The burden of proof is the responsibility to establish those facts that must be proven if the case is to be resolved. The burden of proof falls to the claimant in an adjudication case and to the plaintiff in tort litigation. In principle, the claimant or the plaintiff is challenging the status quo and must provide the information necessary to establish his or her claim and upset the status quo. This means that the defendant shoulders less of a burden, in theory, because it is only necessary to rebut the claim or assertion, not to establish beyond doubt the absence of culpability or responsibility. In practice, the defendant also usually has substantially greater resources and essential information that the plaintiff lacks. Likewise, in an adjudication proceeding, the insurance carrier or agency receiving the claim normally has much greater access to expertise than the claimant. Such cases therefore often end up with an agency or defendant protecting its position with amassed available evidence on the one hand against an incompletely documented claim of damage on the other. Such claims are often supported by facts of the case that establish the potential for harm but that leave a large degree of uncertainty as to the damage that actually took place in that individual case.

**Table 23.1.** Standards of Evidence*Strength of Evidence*

- Relates to burden of proof
- A statement on the means by which an event occurs and the uncertainty of knowing
- Hierarchy (items above the line are persuasive; items below the line are not)
  - Empirical evidence
  - Inferred evidence
  - Informed opinion
  - Speculation

*Preponderance of Evidence (Frequentist)*

- Interpretation of a causal association
- A statement of whether the model suggests an association in fact, taking into account the uncertainty
- Hierarchy (items above the line are persuasive; items below the line are not)
  - Presumptive
  - Probable—"more likely than not"
  - Most probable among causes
  - Possible

*Standard of Persuasion/Belief (Subjectivist)*

- Degree of uncertainty associated with a specific conclusion
- Operates for both witness and judge/adjudicator/jury
- Hierarchy (items above the line are persuasive; items below the line are not)
  - Certain "beyond reasonable doubt"
  - Probable
  - "Reasonable medical certainty"
  - Possible

The burden of proof cannot be described by a statement of certainty, because it is a statement of fact, not supposition. Whether it has been met can be described by a statement about the strength of the evidence. Individual elements of evidence may be compared

and assessed for strength and weight. Table 23.1 also presents a hierarchy of the strength of evidence used to discharge the burden of proof; the higher in the hierarchy, the stronger the type of evidence. Informed speculation, permitted in the post-*Frye*, pre-*Daubert* era, is no longer acceptable; it was always the weakest level of evidence. Informed opinion is stronger because the opinion can be based on statements of uncertainty and general principles with more rigor than speculation. However, inferred evidence (drawn from studies that relate to but are not specific to the individual case) and empirical evidence (actual data from pertinent studies) are much stronger.

Whichever the situation, what ultimately matters is the preponderance of evidence when the evidence on both sides is reviewed. If the evidence on both sides is weak but marginally stronger on one side than another, that satisfies the standard of “more likely than not.” Adjudicators and judges may be unsure just how important a particular bit of evidence may be, and this may be the deciding factor in the case. If the evidence on the two sides is extensively documented beyond dispute and is in every respect (within reason) equally matched, then a wispy bit of data on one side that has not been effectively refuted may tip the balance. Courts take into account the strength of the evidence, but it is the weight that counts most.

The preponderance of evidence is described by statements of certainty. These statements of probability have been called “frequentist probability” because they refer to the relative frequency of events in a sequence of experiments or, in epidemiology, a time period of risk. The corresponding hierarchy of certainty for the preponderance of evidence is also presented in Table 23.1. Terminology is crucially important here. “Possible” is taken to mean less than probable, therefore less than 50 percent certain and therefore not likely. The word “possible” therefore implies that the evidence does not favor a causal role for the factor involved, even if it admits that there may be an association. A risk factor that increases a person’s risk some of the time or only a little for an outcome that is relatively common in daily life cannot be said to be the probable cause, but may be a possible cause among others. This does not establish causation but may carry

some weight in certain circumstances. For example, if it can be demonstrated that the preponderance of evidence favors the cause as a contributing factor but not the sole or principal cause, then the possible cause of the condition may become a probable contributing factor. Under some adjudication systems, this may be enough to establish the claim. More generally, however, the most likely exposure among possible causes is the first solid level of certainty likely to be persuasive, and “probable”—more likely than not—is the general standard for certainty. Where an association is well established as the primary and most likely cause in the great majority of cases, it is sometimes still adopted as a “presumption” by law or policy, as described below; that is, the association is assumed to be causal unless demonstrated otherwise through rebuttal.

The standard of persuasion or belief is the degree of confidence imparted to the adjudicator or jury and internalized by them in making their decision. It is a psychological factor for all parties to the case, including the expert witness him- or herself, but it plays a critical role for the judge, jury, or adjudicator. It is a statement about the security of supposition in the face of the unknowable. It reflects the effort required to overcome the subjective uncertainty the adjudicator may feel over the decision and over not knowing what really happened. This statement of certainty has been called “subjectivist” by Gold because it reflects subjective comfort in arriving at the conclusion.

Table 23.1 also presents a hierarchy of certainty for the standard of persuasion. For the adjudicator to believe that it is “possible” that the putative cause resulted in the outcome is not enough. For this reason alone, highly speculative lines of argument and extraneous facts are usually counterproductive. They often convey a sense of weakness of argument that undermines confidence in stronger evidence, and they cloud the perception of adjudicators that there is one clear answer to the problem. A conclusion with “reasonable medical certainty” implies a considered judgment that the relationship is probable but that there is a range of uncertainty around the judgment that may include the possibility that the relationship is not causal. In the

absence of definitive evidence, the burden of decision-making placed on the adjudicator is increased. The expert witness, after exercising informed judgment, cannot say for sure, although he or she is inclined on balance to accept the relationship as causal. The adjudicator, with less technical insight, must now apply lay judgment a second time (having first decided whether to accept the expert witness as authoritative), in effect choosing somewhat arbitrarily where he or she will fall within the range of uncertainty implied by the expert witness. “Probable” is not only much stronger, but use of the term implies that the range of uncertainty is much narrower, giving the adjudicator much less room to apply personal judgment (assuming that he or she accepts the expert witness’s formulation). Certainty “beyond reasonable doubt” approaches the much higher standard used in criminal cases and is uncommon in these cases.

Gold has pointed out that the frequentist and subjective probabilities are often conflated into a general statement of likelihood. He points out that this confusion distorts the proper analysis of the case in several ways. It devalues the preponderance of evidence because it confuses the notions of weight of evidence and the most convincing evidence, or that evidence most likely to make an impression. It makes it difficult to sort through complicated chains of logic, each step of which is associated with a conditional probability in which the “more likely than not” standard must either apply to each sequential step in the logic (a frequentist approach, relaxing the expected rigor of the testimony) or to an overall summation of the argument (a subjectivist approach, increasing the standard for accepting the testimony). Collapsing also tends to favor certain types of evidence, such as epidemiological studies, that yield clear statements of probabilities and estimates, even though the standard for resolving medical disputes is a conclusion about an individual case, which may be exceptional or for which there may be particular circumstances. It also tends to focus attention on a particular estimate or probability, even though a range or set of values may be more appropriate to the case. Finally, Gold points out that decisions are made on the basis of persuasion—the level of subjectivist probability—not by the other

levels of certainty. These are of course critical in arriving at the final level of persuasion, but it is ultimately the confidence in the minds of the adjudicators that decides the case. That depends on the presentation of the case and the expert's explanation of the significance of the evidence.

## THE ROLE OF EPIDEMIOLOGY

Epidemiology has become one of the most powerful tools in modern tort litigation and adjudication for occupational and environmental hazards. The epidemiologist is now an essential expert for the resolution of many issues involving health outcomes following exposure to hazards. However, epidemiology is not enough.

Epidemiology is fundamentally a science of generalizations. The basic approach of epidemiology to estimating risk is to measure the experience of a population of individuals with the expectation that, all other things being equal, the overall risk for the group will be a valid estimate for most members of the group. Epidemiology has become increasingly valued in health-related cases precisely because it is a powerful tool for generalization.

However, epidemiology has limitations precisely because it is a science of generalizations. That is its great strength, but also its great weakness. When applied to class actions, generalizations make sense because one is considering patterns in a large population. However, most litigation involves individual plaintiffs, and the individual circumstances of the case must be separately considered. (This is also true in adjudication systems, such as workers' compensation; most legislative acts require that individual consideration be given to each claim.) Thus, epidemiology can inform the expert witness with a description of what happens most of the time or what is most probable, but the interpretation still must be brought to the level of the individual case. This may mean demonstrating that the plaintiff or claimant is similar to a group at demonstrably high risk; or that he or she is different and therefore belongs to a subgroup; or that he or she has unique characteristics, so the risk is not adequately described by

summary statistics. This is where a well-prepared, knowledgeable medical expert can play a critical role.

Epidemiological studies have the great advantage of reflecting human experience and the great disadvantage of being generalizations. The science of epidemiology describes patterns in populations, and from this, one infers the most likely history or the most likely future risk in an individual case. However, studies on populations are not the same as predictions for individuals. Individuals may or may not conform to the pattern of characteristics in a population or a relevant subgroup. Individuals have life histories and risk factors that exist in combinations in one person but as frequencies in groups, scattered among individuals in a population. Thus, it is very important to separate three roles of epidemiology as a science in providing evidence for a causal association:

- Observational data that relate the risk factor (putative cause) and effect and provide evidence on a statistical basis from population-based science for causation, even though causation cannot formally be proven by epidemiology
- Frequentist predictions for individuals, based on an estimate derived from the experience of a population from which they were drawn or that they are presumed to resemble
- Population studies that can be used to attribute (but not apportion) cause in a large group, demonstrating that an outcome is associated with a particular occupation, hazard, or exposure
- Descriptive studies that put a problem in context, such as the number of people exposed to the hazard

Causation is an especially elusive concept applied to epidemiology. Human populations are vastly more complex than experimental systems, subject to numerous influences on health, behavior, and social adjustment. Each individual in the population may be subject to numerous other exposures that influence the outcome of interest. Although these are called “confounding” factors, they are often every

bit as important in determining the outcome as the risk factor under study; it is a mistake to dismiss them as merely sources of bias. In the presence of numerous “confounding” factors, clear associations demonstrated in epidemiological studies are remarkable observations, with three likely explanations: an association, a statistical chance event, or a bias in the study method. If the latter two “false” outcomes are excluded, the association that remains is not necessarily causal in nature. Often, the Hill criteria (explained in Chapter 2) are applied in such cases and have become popular in courtroom testimony. However, this is incorrect. The Hill criteria are general guidance within the framework of epidemiology and are neither definitive nor grounded in the weight of evidence.

For the epidemiological science to be brought down to the specific case, the science must be understood and then interpreted with knowledge of the circumstances of the specific case. The medical expert contributes most to understanding in going from the broad generalization, which is called “general causation” in American law (does  $x$  make  $y$  happen?), to an opinion on causation in the particular case (did  $x$  make  $y$  happen in this individual?), which is called “special causation.” This requires an expert who understands the epidemiology and also understands the individual. OEM physicians spend their entire careers doing just this.

Another critical issue is how to apply scientific evidence when the standard is “more likely than not,” rather than scientific certainty. In other words, what would epidemiology and biomedical sciences be like if the standard for conclusions (not necessarily individual experiments) were 50 percent rather than 95 percent certainty? What would be the role of the doctrine of “falsification” (the notion that a theory cannot be proven, only disproven) if the standard for accepting a theory were only the weight of evidence rather than a single fact that the theory cannot explain? In effect, this is the issue that confronts the expert in developing a theory of what happened in the individual case, also known as “framing.”

Here, epidemiology faces a serious methodological limitation that it has not acknowledged within the discipline. Inference based on

conventional “frequentist” statistics are based on an a priori probability: what are the chances of something happening in the future, extrapolated from what happened in the past? An a priori risk is meaningful only before the event has occurred. After the event has occurred, its occurrence is no longer a prediction, and its causes must be sorted out; the risks must be put into the context of an individual risk profile, which requires knowledge of the “posterior” probabilities, or the probability of an association in the past once the even has occurred, however unlikely it may have been going forward. There are some experts who believe that conventional statistics cannot be applied in such cases and that Bayesian statistics are more appropriate, as they are in determining diagnosis in a clinical case. Perhaps because of their complexity, novelty in court, and how difficult they are to explain to nonstatisticians, this is never done in practice. It is at this point that the argument for medical dispute resolution should be based on individual evidence rather than epidemiology alone. Because Bayesian statistics are simply not used, notwithstanding the logic for applying them, epidemiological evidence is always of the conventional “frequentist” variety. (See Chapter 21 for a similar argument applied to workers’ compensation.)

The frequentist point of view requires that the epidemiologist place him- or herself in a position in the past, before the event occurred, and to determine whether the probability of the illness or injury was high or low before it occurred. The “bright line” in that construction is whether the risk estimate (relative risk, standardized mortality ratio, or odds ratio) is consistent with a doubling of risk in absolute terms, because a doubling of risk corresponds to the weight of evidence for one individual in the population, all other things being equal. A doubling of risk means both that it is more likely than not that a particular cause or hazard was associated with the outcome in any one case, and also that half the cases are associated with this hazard or cause. It can therefore be applied to the individual or to the characterization of group risks, and therefore it confers eligibility for a “presumption” (the policy that a given case with a given association will automatically be accepted as causal unless there is good reason

to think otherwise). For any one case arising out of a group of, say, workers with a risk that is doubled or more (for example, a case of kidney cancer arising in a firefighter), it can be presumed that unless the case had unusual features, the condition was as or more likely than not to have arisen from the cause or hazard. In workers' compensation, presumption is an important way of simplifying procedures and recognizing risk among occupations and workers exposed to particular hazards. In civil litigation, doubling of risk is strong evidence that the association is causal, because it is a relatively strong association and explains the outcome.

The rationale behind making a doubling in risk a criterion for presumption requires some explanation if it is not already obvious. The standard of an exact balance of probabilities means, in this context, that the risk of developing a bad outcome arising from the hazard is equal to the risk of the outcome arising from other causes, or for simplicity, in the general population. In statistical terms, therefore, the attributable risk from the hazard is equal to the expected risk. The balance of probabilities also means at least equal odds, or 1:1, or an odds ratio of 1. A balance of probabilities also means that the overall probability that the condition was due to the exposure or putative causal factor in the population under study is 50 percent, and one half equals the other half. A relative risk of 2 or an SMR of 200 means that the risk due to the exposure is equal to the risk of the general or comparison population. (If the risk of the general population is 1, then a relative risk of 2 can be thought of as 1 plus 1 divided by 1.) This in turn corresponds to an SMR of 200, an odds ratio of 2, or an attributable risk fraction 100 percent of expected. Therefore, if an epidemiological study has a relative risk of 2, then all other things being equal, it is strong evidence for the balance of probabilities and favors the conclusion that there is a causal association by legal standards of civil litigation, although not by the standards of science.

In practice, the criterion of a doubling of risk can be liberally applied, as it should be, given the uncertainties and the wide confidence intervals surrounding most of the available estimates. This

criterion still excludes many disorders for which there is a clear association but for which nonoccupational risk factors appear to be much more important in determining risk.

## PRESUMPTION

Modern presumption developed as a simplifying doctrine within the workers' compensation system. The concept of presumption makes sense when a disorder is well known—and well documented—to be strongly associated with particular workplace or environmental conditions. A claimant who presents with a particular diagnosis or type of injury in the particular setting that is characteristic for the disorder is presumed to be eligible for compensation and benefits, as long as the conditions are documented.

The justification for presumption is that the disease is either so common within a particular occupation or so rarely found outside certain occupations that any given claimant almost certainly developed the disorder as a result of work-related exposure. The simplest example is mesothelioma, which, in the presence of a work history suggestive of exposure to asbestos, is almost universally accepted as work related. The most controversial presumptions are those for elevations in risk of common disorders. This is the general problem with presumption in the case of most adverse health outcomes relevant to occupations such as firefighting, where lung cancer has been a persistent issue.

In practice, presumption takes one of three forms:

- Scheduled or “prescribed” occupational diseases
- Rebuttable presumptions
- Legislated entitlement programs

Scheduled or prescribed occupational diseases are accepted without any questioning of their association with exposure. Much of the legislation proposed in the middle of the first decade of this century

for solving asbestos litigation scheduled asbestos-related diseases. Schedules are also common in countries other than the United States and Canada.

A rebuttable presumption is a policy that a disease will be considered as arising out of work unless there is a reason to dispute (rebut) this conclusion. This is the most common type of presumption in the United States and Canada. The rebuttal is based on the circumstances of the case and the characteristics of the individual worker. The general association between the disorder and occupation is not itself disputed, only the association in the particular case. Although a rebuttable presumption may be contested by the employer, the burden of proof is on the insurance carrier or the employer to demonstrate why the claim should not be accepted. Many rebuttable presumptions are passed as legislation: The states of California and Virginia have enacted rebuttable presumptions for chronic obstructive pulmonary disease or cancer for firefighters, and for myocardial infarction for firefighters and police.

Legislated entitlement programs are presumptions that have been written into law—for example, the Black Lung Benefits Program.

Ideally, rebuttable presumptions should be limited to disorders in which the group risk is at least doubled, such that in every case, the odds are as or more likely than not the result of work-related factors. However, legislated schedules are often based on notions of risk that cannot be documented scientifically or that are historical, but not necessarily current. Once legally established, it is almost impossible to remove such provisions.

Presumption rests on a set of assumptions that are rarely examined in detail. A general presumption of risk rests on the conclusion that any given claim from a particular group (such as workers exposed to lead) for a particular disease (such as lead poisoning) that comes before the adjudication agency is (much) more likely than not to be work related. The claim can be expedited on that assumption. Presumption requires that the individual case match circumstances known to give rise to the condition and that the case is drawn from a specific, defined population known to be at risk. As a practical

matter, the probability that the injury came about because of any other cause must be so low that assessing causation is not worth the administrative effort.

This level of certainty may be achieved in two ways:

- Exclusive causation (e.g., pneumoconioses)
- General presumption of risk

Exclusive causation is easy: There are few nonoccupational causes of some diseases (such as silicosis), and the question is only where and when exposure occurred.

General presumption of risk is much more problematic. The analysis of presumptive occupational causation in a disease that is not unique to occupational exposures is a process with multiple steps. First, the association with occupation or, preferably, a specific occupational exposure must be demonstrated. Usually, this is done by means of epidemiological studies, from which a population-specific relative risk (called “population risk”) is derived. Finally, the criteria have to be shown to apply in the individual case. Then, rebuttals are formulated and presented or anticipated. These actual hypothetical rebuttals are addressed in the report on file to ensure that other causes have not been overlooked.

Presumption must satisfy the usual compensation standard of proof that the occupational cause must be “more likely than not” responsible for the outcome (giving the benefit of the doubt to the claimant). Therefore, a general presumption of risk requires demonstration that the risk associated with occupation must be at least as great as the risk in the general population. This can be shown if the usual measure of risk in epidemiological studies is at least double the expected risk, making allowances for uncertainty in the estimate. Therefore, a general presumption of risk is only justified if there is a consistent and plausible association with occupation, and the magnitude of risk is at least double that of the general population.

General presumption of risk, on the basis of population estimates, assumes that all members of the population or occupation share in the pooled risk about equally or randomly, and that it is more likely than not, in the individual case, that the cause of the condition is work related. In terms of epidemiology, presumption reflects an attributable risk that is greater than the expected risk in an unexposed or general population. One cannot know for certain that any individual case drawn from the population is due to this exposure, but the odds are enough that it is a reasonable working assumption. A general presumption of risk is not the same as identifying an association, which can be any magnitude. A general presumption of risk implies that the elevated risk is sufficient in magnitude to make it more likely than not that in the individual case the cause of the disorder was related to work. A presumption means that there is not only a relationship, but that it is strong enough and consistent enough to be causal, and that it accounts for more than half of the cases of the disease. Both assume that some association with the exposure or occupation has already been proven to a reasonable standard of scientific or legal proof.

When population risk data are applied to individual cases, there are two questions to be considered in evaluating the association between a putatively work-related condition and occupational exposure: (1) How strong is the evidence for an association? and (2) How strong is the association itself? The Bradford Hill criteria (outlined in Chapter 2) address the former but not the latter. For population risk data to be useful as applied to the individual, however, some indication must be given that the individual risk, inferred from the strength of the association in the population, is at least as great as that from other causes not associated with occupation. This criterion must be satisfied before concluding that an occupational cause is likely to be causal in an individual case. The doubling criterion for epidemiological data does this on a population basis and makes it much easier to argue causation in the individual case. Its value is in evaluating a general presumption that all (or the great majority of) workers in a given occupation are at risk. The specific case must still be evaluated

individually. When a general presumption of risk cannot be made, individual assessment of risk should be used to deal with the merits of the case. Some observers believe that in some jurisdictions, such as Texas, risk doubling has been elevated by judicial decision to the exclusive basis for accepting causation in civil litigation. This may be going too far, because epidemiological findings have too much inherent uncertainty, potential for dilution of risk estimates, and potential for bias to be so doctrinaire.

A general presumption of risk is not easily justified or defended for weak associations or when diseases are common in the general population. A more productive approach may be to take the claims on a case-by-case basis, examining individual risk factors and overall risk profile. A general presumption of risk is more easily applied to unusual disorders with high relative risks, particularly when they are unique to or characteristic of certain occupations.

The expert must apply the population risk to the individual case as guidance, not as an accurate estimate of personal risk. A population risk, which conforms to the past experience of the population under study, is only an estimate of the individual risk. Characteristics of the individual—such as age, length of employment, family history, and lifestyle—may substantially modify the individual risk profile. For an individual with few risk factors for the disorder, the population risk may be an overestimate of total risk but an underestimate of the risk due to occupation. Conversely, for an individual with many nonoccupational risk factors for the disorder (or occupational risk factors unrelated to a claim), the population risk may be an underestimate of total risk but an overestimate of the risk that can be attributed to occupation. Rebuttal rests on assessing personal characteristics on a case-by-case basis.

## **WITNESSCRAFT**

There are skills to serving as an expert witness, just as there are skills a physician must master as a clinician. These skills suffer neglect when the expert becomes absorbed in on his or her own presentation skills

and value to the case. They flourish when the expert puts ego aside and concentrates on the case.

The responsible expert should be prepared and should know the relevant literature. This does not mean stopping when one finds a passage or evidence supporting a conclusion. This means reading beyond the literature that supports one's position, knowing the arguments on the other side, and being aware of controversies that affect the conclusion.

There is a natural tendency for highly educated people to lapse into the language of their discipline because it is precise, comfortable, and establishes authority. However, it is a destructive tendency that must be resisted in testimony and reports. The expert must cultivate the skill of describing things in common words, explaining complicated things simply, and translating jargon into everyday language. This also means knowing exactly what the jargon means (it can be very embarrassing to have one's definition of a medical term questioned by a knowledgeable lawyer).

The best experts also have a sense of how technical terms are actually used in conversation between experts, as well as how the words would be defined, formally and in common language. For example, the word "neoplasm" is formally defined as a new tissue growth that may be benign or cancerous, but the term is almost never used in that sense outside of pathology textbooks. The specific word "neoplasm" is almost never used between physicians in conversation except as a euphemism for suspected cancer; when that word is used, usually with a pause, it strongly implies that one physician is expressing to the other, if only subliminally, that he or she is worried that a patient's mass is indeed cancerous despite indeterminate tests, and it implies that the physician is trying to figure out how to manage the expectations of the patient in the face of uncertainty.

The expert examines the internal evidence of the case, paying particular attention to anomalies and unexplained facts, which might be the key to alternative explanations. The expert places the case in the context of the literature and external, mostly scientific, evidence, demonstrating how the case fits the profile described by the literature

or how the individual case is sufficiently different that the literature does not apply.

The expert will be expected to produce his or her current CV. This should be complete, with explanations entered for all lapses in credentials, periods out of practice, or major career changes, particularly if they are unusual. Unexplained gaps or discontinuities in practice invite scrutiny and challenge at the time of deposition. The expert will need to supply a list of cases in which he or she served as a witness in recent years, indicating which were on the side of the plaintiff and which were for the defense. (The time period varies with the jurisdiction.) If testimony in recent cases has all been on one side or the other, the opposite side will take it as evidence of bias; and if the testimony has been balanced, the opposite side will insinuate that the expert's opinion is for hire. That is how the game is played.

Discovery is the process by which all the records and documentation held by one side are made available to the other. It occurs before trial and is intended to level the playing field so that, as much as possible, the factual evidence can be agreed upon and there will be no surprises. The expert's notes and records are "discoverable" and should also be produced at the time of deposition and trial. These records include marginal notes, personal comments, sticky notes, handwritten outlines, interim and rough drafts, and analyses that may have been preliminary and possibly incorrect at the time. If they show up in a discovery, they will be legitimate subjects for inquiry, and the most embarrassing or inappropriate will certainly be flaunted by the opposing counsel. The result is usually to make something irrelevant to the case into a major issue prejudicial to the expert. Experts should keep their files free of this extraneous material, refrain from making marginal notes, never make rough drafts, and never write on any document that is discoverable.

Bills and invoices should be kept, but in a discoverable file separate from the records of the case. The expert will be asked to disclose his or her rate, and opposing counsel will usually imply that it is too high and that the expert has been bought. An extremely high fee tends to

work against the credibility of an expert. However, an extremely low fee raises eyebrows (unless the expert is working “pro bono”—donating his or her time).

The expert must keep a record of key documents, sources, and reference materials on which the expert “relied” in coming to a decision. Most experts base at least part of their opinion on common knowledge in the field. That common knowledge is very likely to be questioned if a specific, relevant textbook passage or authoritative source is not added to the list. Courts are not very tolerant of long lists of standard textbooks or general references: they expect specific articles, chapters, and page numbers. The expert should make copies of all documentation “relied upon,” because the lawyer on one side is often required to furnish copies of the relevant textbook pages or articles to the opposing side.

The lawyer will ask for time to “prepare” the expert before trial or deposition. This preparation time is immensely valuable and should not be rushed. The expert is usually focused on a particular technical issue or theory of causation. Because the expert is rarely involved with a case every day for days on end, it is easy to lose track of the sequence of events, names and events that are pivotal in other aspects of the case, and the meaning of the hearing within the overall process (for example, if it is to try the case, to qualify a class action, or to establish admissibility of evidence). The lawyer, however, lives with the case and should, by the time of deposition and trial, have a better knowledge of the facts of the case than the expert. The lawyer will refresh the expert’s memory on many forgotten aspects of the case that may come up. The lawyer will want to hear, one last time, what the expert plans to say and how he or she will say it. This is also a time when the lawyer will impress upon the expert the particular legal burden of proof in the case and terminology that has special meaning in that jurisdiction.

Depositions are pretrial question sessions in which lawyers examine and cross-examine experts and other witnesses on the record. They will be available to the judge and the evidence provided in deposition may be used later in trial. The purpose of the expert’s

deposition is to discover the expert's argument, command of the evidence, and theory of the case before the trial. It is an important step in the case, but only rarely is testimony given in deposition later discussed at the trial itself, and then only when it is flagrantly contradictory or when it appears that new testimony is being introduced at the trial that was not covered in discovery and deposition. Deposition testimony is often important in signaling to the opposing side the strength of the case and motivating them to settle.

There is a predictable pattern to questioning on the stand and in depositions. Counsel for the side that has called the expert will usually start first, and one of the first items covered will be the qualifications of the expert. The expert should show neither arrogance nor defensiveness; no matter how much emphasis the attorney places on them, qualifications are only to set the stage and justify why the expert has been called. Opposing counsel, starting cross-examination, will almost always start off with a pleasant and cordial manner and then turn aggressive and even hostile, often feigning outrage over some contentious point; one should always expect it. The expert should answer the question and refrain from giving a seminar on the topic. This is a situation in which less is often more, and a long-winded answer carries a risk of raising secondary issues, distracting from the pivotal issues, or creating misunderstandings that do not help the court. This is no time to deliver a sermon. The expert should not volunteer information. A good lawyer will already know the answer to every question he or she will ask and will adapt and change tactics in the event of an unanticipated response, leaving the expert one step behind. Experts who try to outsmart the lawyer often end up confusing and outsmarting themselves. "Trick" questions are perfectly acceptable in the courtroom, and the expert who cannot handle them, unpack their various parts, and explain the answers on the levels required probably should not be on the stand. The expert should take his or her time, answer thoughtfully, and stop when the question is answered without going further and introducing new and confusing elements into the discussion. If the deposition is being recorded on video, however, overlong pauses may seem like equivocation.

The trial is similar to the deposition in format, but it is a very different experience. All eyes will be on the witness. In the courtroom, unlike in a deposition, hesitation in giving answers gives the impression of insecurity and sometimes falsehood. The expert is bound to stick to what has been covered in the deposition and cannot introduce entirely new lines of argument or a new and undiscovered pivotal fact. At trial, it is never entirely clear whether the jury, or for that matter the judge, has actually understood the testimony. In general, juries seem to understand more than they let on, and trying to "read" a jury is hard enough for a lawyer and pointless for an expert.

Aids to testimony are distracting and should be kept to a minimum unless they make a major difference in understanding. Models, graphics, PowerPoint presentations, and other aids interrupt the flow of testimony and are often interpreted too literally by jury members, who may get hung up on a detail and miss the bigger picture. They are most useful in showing anatomical relationships, spatial relationships, and details of key pieces of evidence. They should be used by the lawyer, not the witness.

Lawyers are at least as smart as experts, and many of them are much smarter. Experts should never assume that they are smarter than the lawyer. One should not try to play intellectual games in his or her testimony, nor should one try to engage in one-upmanship or try to anticipate where a lawyer's line of questioning is going.

It is an accepted tactic for the opposing counsel to attack the qualifications, credibility, honesty, motivation, and competence of the expert. These attacks can be stinging but should not be confused with personally motivated attacks. Their only purpose is to impugn the evidence, and one easy but intellectually bankrupt way to do this is to show that the expert is undisciplined and untrustworthy. Baiting the expert in order to cause him or her to become angry or defensive is a sign that the opposing counsel has run out of steam and is frustrated in its efforts to counter the evidence. It is a desperation maneuver, especially at trial, because the lawyer risks appearing abusive to the jury. It is essential that the expert react not by losing

control, but by remaining professional and therefore credible throughout the personal attack. It should be viewed as a backhanded compliment to the preparation of the case. It is not about the expert.

It is also important to realize that the counsel representing the side the expert is assisting is motivated to win the case, not to make the expert look good. The lawyer has a strategy and knows, or should know, how he or she plans to use the expert. The expert should have no strategy, because the case is not about him or her.

New experts are often concerned about what they should wear and how they will appear on the stand. This is a very minor part of being an expert witness. An expert does not have time to establish his or her persona and the depth of his or her understanding of the field in the short time available on the stand. As a result, much depends on first impressions. The expert should therefore look the part of a responsible and meticulous physician: professional, well-groomed, conservative in dress, confident but not arrogant, and careful in manner and use of words. Beyond that, the expert should look as neutral as possible, because anything that draws attention to the expert distracts from the case itself.

In preparation to do this kind of work, the OEM physician should have “errors and omissions” professional liability insurance, which goes beyond clinical practice. It is important to maintain a clinical practice at all times, because current engagement in providing patient care is important in establishing the expert’s qualifications and credibility. Mixing plaintiff and defense work is helpful in establishing credibility, and argues that the expert is responsible and takes cases on their merits. Whether to advertise and list with commercial directories is a personal choice, but many experienced experts prefer not to do so, because it creates the appearance that they are “hired guns.”

## EVIDENCE-BASED MEDICAL DISPUTE RESOLUTION

Medical knowledge has many uses outside of medicine. When operating outside the medical profession, the physician must consciously accept that medical norms do not necessarily apply.

Extension of biomedical research into commercial applications in the pharmaceutical and biotechnology industries is an obvious example. In these settings, the principles of business and engineering run the show, not those of the physician or biomedical investigator. Medical knowledge is also useful to resolve conflicts arising from injury, regulation to prevent future risk, and public policy. When this occurs in the context of the resolution of legal disputes, the physician plays the role of expert, applying medical knowledge to a practical problem that is outside the context of medical practice or biomedical research and that is not governed by the rules of medical practice or research. In court, the judge runs the show, not the physician serving as a medical expert witness. The expert defers to the logic of the legal system.

On a practical level, the legal system lacks the capacity to evaluate the validity of knowledge as evidence and relies heavily on expert opinion. There are no broad, socially agreed-upon rules for the application of medical knowledge, public-health knowledge, or for that matter any scientific knowledge except in the rules of evidence and decisions of the law. This is particularly evident in tort litigation, when liability for causing injury is under consideration, and often rests on theories of disease etiology and the circumstances surrounding exposure to a hazard.

A similar problem once existed for the clinical practice of medicine. Over the last twenty years, an approach called “critical appraisal” has established norms for the acceptance of evidence in clinical practice that are now almost universally accepted. Critical appraisal is a systematic approach to evaluating the evidence based on clinical epidemiology; evidence-based medicine is the practice of medicine justified by valid studies correctly interpreted.

Is it possible to develop a framework for applying the knowledge of health and medicine that is similar to the concept of critical appraisal? How can the evaluation of medical knowledge be adapted to the rules of the dominant framework of dispute resolution in our society: the law? Is it possible to develop an evidence-based approach to the expert’s work, which one might call “evidence-based medical dispute resolution”?

Thirty years ago, a movement toward evidence-based medicine revolutionized clinical practice. Critical appraisal of the medical literature and the reliance upon evidence-based principles by managed-care organizations and utilization-review organizations led to the adoption of evidence-based medicine as the dominant mode of clinical practice today. The concept of critical appraisal and evidence-based medicine was not embodied in legislation or enforced as governmental or judicial policy. This movement advanced for many years through education in medical schools, debate, and consensus until it was ready to be institutionalized in practice. It became the accepted norm because it met a need, satisfied a rising demand, and made sense to all participants. Evidence-based medicine did not end controversy in medical practice, but it confined the scope to the scientific issues and rooted controversy in evidence rather than unsubstantiated opinion.

The current state of affairs in the courts is not unlike the situation in medicine at the time clinical epidemiology was “invented.” The “practice” of medical expert witnesses is not standardized or governed by a consistent set of principles. Each expert witness is essentially autonomous. An expert witness cannot link, at present, to a community of other experts who have a consistent view of how to approach a problem or interpretation. Medical practitioners thirty years ago were similarly autonomous. That all changed for medical practitioners as a result of increasing external demands for consistency and persuasion.

It should be possible to develop a similar framework for the evaluation of scientific evidence in legal settings. It will not be possible—or even desirable—to distill a set of rigid rules for dealing with scientific evidence in legal settings. However, if the broad outlines of reasonable interpretation can be agreed upon, all parties will have advanced much further and can concentrate on the factors of the individual case. There is a need for evidence-based medical dispute resolution.

There is no formula or easy set of rules that can be derived for the universal application of this approach to scientific evidence. What would evidence-based medical dispute resolution consist of? A

rational approach to evaluating evidence in the health sciences requires both a capacity to generalize, usually on the basis of a population, and a capacity to individualize to the specific case. If the mechanism is known, the explanation enhances the credibility and therefore the persuasiveness of the conclusion. This approach should be useful in the development of a specific case and in guiding the development of the administrative systems in which it is used. Ideally, it should contain these elements:

- Epidemiology and the interpretation of population data
- Individualization of the evidence to the specific case, using methods of clinical medicine, toxicology, and (in the future) genetics
- Statistical treatment that does not necessarily rely on conventional assumptions designed for scientific studies
- An understanding of science that takes into account the social nature of the scientific enterprise, as shown in contemporary studies in the history and philosophy of science
- Adaptability to a variety of applications, including public policy, statutory adjudication systems, and tort litigation

## RESOURCES

- Gold S. Causation in toxic torts: burdens of proof, standards of persuasion, and statistical evidence. *Yale Law J.* Dec 1986;96:376–402.
- Guidotti TL, Rose SG, eds. *Science on the Witness Stand: Evaluating Scientific Evidence in Law, Adjudication and Policy*. Beverly Farms, MA: OEM Health Information; 2001.
- Jasanoff S. *Science at the Bar: Law, Science and Technology in America*. Cambridge, MA: Harvard University Press, Twentieth Century Fund; 1995.
- Meufeld PJ, Colman N. When science takes the witness stand. *Sci Am.* 1990;262(5):46–53.
- Muscat JE, Huncharek MS. Causation and disease: biomedical science in toxic tort litigation. *J Occup Med.* 1989;31(12):997–1002. [This paper explains the importance of a doubling of risk and why it is equivalent to the “balance of probabilities.”]

## NOTEWORTHY READINGS

- Black B. Evolving legal standards for the admissibility of scientific evidence. *Science*. 1988;239:1508–1512.
- Black B, Lilienfeld DE. Epidemiologic proof in toxic tort litigation. *Fordham Law Rev*. 1984;52:732–785.
- Brent R. Medical, social, and legal implications of treating nausea and vomiting of pregnancy. *Am J Obstet Gynecol*. 2002;186(suppl 5):S262–S266. [A particularistic discussion of the *Daubert* case.]
- Cohen FL. The expert witness in legal perspective. *J Legal Med*. 2004;25(2):185–209.
- Goss PJ, Worthington DL, Stallard MJ, Price JM. Clearing away the junk: court-appointed experts, scientifically-marginal evidence, and the silicone breast implant litigation. *Food Drug Law J*. 2001;56(2):2727–240.
- Guidelines for the physician expert witness. American College of Physicians. *Ann Intern Med*. 1990;113(10):789.
- Guidotti TL. Evidence-based medical dispute resolution in workers' compensation. *Occup Med*. 1998;13(2):289–302.
- Guidotti TL. Evidence-based medical dispute resolution. In: Demeter S, Andersson GBJ, eds. *Disability Evaluation*. 2nd ed. Chicago: American Medical Association; 2003:86–94.
- Guthell TG, Hauser M, White MS, Spruiell G, Strasburger LH. "The whole truth" versus "the admissible truth": an ethics dilemma for expert witnesses. *J Am Acad Psychiatry Law*. 2003;31(4):428–431.
- Hollingsworth JG, Lasker EG. The case against differential diagnosis: *Daubert*, medical causation testimony, and the scientific method. *J Health Law*. 2004;37(1):85–111.
- Kulich RJ, Driscoll J, Prescott JC Jr, et al. The *Daubert* standard, a primer for pain specialists. *Pain Med*. 2003;4(1):75–80.
- Meyer C, ed. *Expert Witnessing: Explaining and Understanding Science*. Boca Raton, FL: CRC Press; 1999.
- Price JM, Rosenberg ES. The war against junk science: the use of expert panels in complex medical-legal scientific litigation. *Biomaterials*. 1998;19(16):1425–1432.
- Price JM, Rosenberg ES. The silicone gel breast implant controversy: the rise of expert panels and the fall of junk science. *J R Soc Med*. 2000;93(1):31–34.
- Reference Manual on Scientific Evidence. 2nd ed. Washington, DC: Federal Judicial Center; 2000. [Use current edition.]
- Science, Technology and the Law. Science and Technology Policy Forum. New York: New York Academy of Sciences, Science in Society Policy Report; 1998

# **24 REGULATORY AGENCIES**

This chapter provides an overview of the principal regulatory agencies in the United States and a brief description of the much less complicated but province-specific regulatory framework in Canada. The regulatory frameworks for specific hazards, workplaces, special occupations, and environmental risk are discussed where these respective issues are covered elsewhere in the book.

The occupational and environmental medicine (OEM) physician should know the missions of these agencies as they pertain to both environmental risks and occupational health protection. However, this chapter is for general orientation only. Specific guidance from legal counsel or another authoritative and competent legal source should be sought for any question of law or for any question regarding jurisdiction of these agencies, particularly as it may apply to a particular issue. Do not rely on this or any other text or general reference book for legal guidance.

## **ENVIRONMENTAL PROTECTION AGENCY**

The National Environmental Protection Act (NEPA) of 1970 created the Environmental Protection Agency (EPA), which soon became the largest and arguably most powerful regulatory agency in

the United States government. Almost all significant environmental health protection and some occupational health functions fall within the purview of the EPA. NEPA gave EPA wide authority for environmental protection, including:

- Standards and enforcement for water quality and emissions (pollution release)
- Standards and enforcement for air quality and emissions
- Standards and enforcement for land disposal and hazardous materials
- Standards and enforcement for radiation and radioactive substances in the general environment, other than nuclear waste
- Standards and enforcement for solid-waste disposal
- Standards and enforcement for asbestos in school buildings
- Standards and enforcement for large-scale ecosystem disturbance such as stratospheric ozone depletion and climate change
- Standards, enforcement, and prioritization for abandoned hazardous waste sites, and management responsibility for abandoned sites of high priority
- Standards for pesticide residues in food
- Guidelines for indoor air quality and safe housing
- Protection of ecosystems and adverse effects within its mandate that affect endangered species (the Department of the Interior is the lead agency for this, however)
- Protection of oceans and wetlands
- Planning for and response to oil spills
- Registration and regulation of pesticides
- Occupational health protection for pesticide workers
- Occupational health protection for hazardous materials (hazmat) workers

- Registration and regulation of toxic substances
- Enforcement of standards and regulations under the applicable legislation
- Grant-making authority for environmental-protection infrastructure
- Research on environmental problems, both intramural (conducted by scientists within the agency) and extramural (funded by grants and contracts)
- Environmental monitoring and dissemination of information
- Reviewing and ability to deny approval of Environmental Impact Assessments, a comprehensive review of the environmental implications of any federally supported project
- Reviewing and ability to approve state implementation plans for air and water pollution control
- Promotion of energy efficiency consistent with the Energy Policy Act of 2005 by using its authority to stimulate use of renewable energy, mandate fuel efficiency and blending, and prevent greenhouse gas emission
- Management of the implementation by the United States of treaties and international conventions related to environmental protection
- Programs in support of environmental justice, children's environmental health, emergency planning, and community right to know

The EPA also has oversight authority over state, territorial, and tribal departments of environmental protection. With few exceptions, EPA does not directly regulate utilities, such as water distribution companies, because these are normally supervised directly by the state department of environmental protection. EPA does not handle local environmental issues that fall under state or municipal jurisdiction, such as landfills, although it sets standards and guidelines.

The EPA plays only a supporting role in many important environmental issues that are primarily handled by other federal departments. It is only peripherally involved in the following:

- Noise (local government agencies; EPA once did regulate ambient noise, but no longer)
- Protection of habitat (Department of Interior)
- Endangered species (U.S. Fish and Wildlife Service)
- Nuclear waste (Department of Energy, Office of Civilian Radioactive Waste Management)
- Wetlands management (U.S. Army Corps of Engineers)
- Indoor air quality (Consumer Product Safety Commission for building materials and mobile homes, local public-health agencies for safe housing)
- Traffic-related dust and noise (local government agencies)
- Emerging infections (Unless waterborne, EPA is not usually the lead agency for infectious agents; the Centers for Disease Control and Prevention provide this role.)

The EPA is organized by problem sector and by region. There are ten federal regions, each with a regional administrator and an organization chart similar to the federal EPA's, but proportioned to local issues and including some offices specific for problems, such as protection of the Chesapeake Bay in EPA Region 3 (based in Philadelphia and covering Maryland, Virginia, Delaware, Pennsylvania, and West Virginia). Aside from the staff functions and administrative offices, there are assistant administrators, who each manage several "offices" devoted to major divisions of the environmental mandate. The administrative divisions on the level of the assistant administrators include:

- Air and Radiation
- Enforcement and Compliance

- Environmental Information
- International Affairs
- Prevention, Pesticides, and Toxic Substances (which is mostly concerned with individual or categorical pollutants, and with administration of the Toxic Substances Control Act and pesticide registration)
- Research and Development (including major laboratory complexes in North Carolina, Cincinnati, Las Vegas, and elsewhere)
- Solid Waste and Emergency Response (which is unified by hazardous materials and mainly deals with toxic materials and covers Superfund, brownfields, underground storage tanks, and emergency management, including hazmat).

The EPA is one of the few remaining major regulatory agencies that retains intramural research facilities for work useful in setting and validating standards. This is counter to trend: OSHA is advised in this regard by the National Institute for Occupational Safety and Health; the Nuclear Regulatory Commission by the Department of Energy. However, this legislative trend appears to be reversing because the Food and Drug Administration, faced with a need for much better scientific grounding for its regulatory activities, has massively expanded its internal research infrastructure in recent years.

## **History**

The history of EPA is perhaps best understood through the history of its legislative authority. The U.S. Environmental Protection Agency (EPA) came into being in 1970, the same year as OSHA, after the passage of the National Environmental Protection Act of 1970. A number of environmentally related agencies were detached from various federal agencies and consolidated to create the new EPA, which had the great advantage of bringing enforcement, state grants for environmental protection, standards setting, mission-oriented research, and training and education under a unified management

structure. EPA also combined the authority to protect living species and to conserve ecosystems (mostly a domain of the Department of the Interior) with the authority to protect human health.

The EPA quickly established itself as one of the most powerful agencies in the federal bureaucracy. One reason for the influence EPA has exercised has been the intensity of public concern over environmental issues. The year NEPA was passed, 1970, was also the year of the first Earth Day commemoration (April 22, annually) and the crest of the popular environmental movement that started in the 1960s. Another reason was the considerable legal authority vested in the new agency; it had the power to obtain convictions resulting in jail terms and to block expenditures for huge public projects. However, an often underappreciated source of its influence was its large budget; this derived for the most part from the EPA's role in funding water projects, which in the early history of the agency made up a disproportionately large share of the federal budget for environmental protection and a significant source of funding for public works at the state level.

In 1970 Congress passed the Clean Air Act, which, with the Federal Water Pollution Control Act (which was amended in 1972), gave EPA huge and centralized authority over the two major worlds of pollution control of the day. The Federal Pesticide Control Act (1972) gave EPA authority over individual toxic chemicals, as well as media pollution, and the authority to ban DDT in 1972; this came in addition to the EPA's responsibility for administering the Federal Insecticide, Fungicide, and Rodenticide Act (1947). This authority kept growing with the Toxic Substances Control Act and the Resource Conservation and Recovery Act, which dealt with hazardous materials, both in 1976.

The Clean Water Act, also passed in 1970, gave EPA a mandate to act as a public-health agency that other environmental protection agencies, at the state level and in other countries, have usually lacked. This authority permitted one of the great public-health achievements of the time: the ban on lead in gasoline, which began in 1973 as a gradual phaseout to 1986.

In 1979 EPA introduced a major regulatory innovation known as the "Bubble Policy," which allowed local sources of air pollution to determine for themselves how they would manage their pollution, as

long as their total emissions (within the “bubble”) were in compliance. This innovation gave the agency recognition for regulatory flexibility and enhanced its authority over regulated sources of pollution. It also freed EPA from the need to prescribe specific control measures, which would have been very unpopular and would have constrained technology. Perhaps more important in the short term, however, the Bubble Policy allowed industry to make progress quickly by controlling emissions where it was easiest to do so. This created an early win for the agency and the expectation of further progress as technology improved. That same year, EPA ordered and sued to obtain cleanup at Love Canal, a highly visible toxic-waste disposal site in Niagara Falls, New York, a strategy that served as a model for managing hazardous waste sites.

On the other hand, EPA did not hesitate to ban chemicals outright, as it had DDT, and in 1978 the agency banned fluorocarbons in aerosol products as a measure to protect the stratospheric ozone layer and then banned PCBs in 1979.

In 1980 the agency had its authority expanded even further—and its budget increased by billions—by the passage of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), which created the Superfund for the remediation of hazardous waste disposal sites. Superfund faced its first challenge in 1981 in funding cleanup of the Valley of the Drums, an abandoned site for chemical waste dumping near Louisville, Kentucky.

In the mid-1980s, during the Reagan administration, the agency was criticized for pliant leadership and allegations of political favoritism. The founding administrator, William D. Ruckelshaus, was brought back and reappointed to bring credibility back to the agency; he brought in a new policy of transparency called the “Fishbowl Policy.”

During the 1980s, Superfund, because of its magnitude and complexity, had increasingly come to dominate the agenda of the EPA. By the 1990s, the slow pace of cleanup of abandoned, illegal, or unclaimed hazardous waste sites under the Superfund program was widely seen as an embarrassment to the agency. However, there was

progress on many other issues, notably air toxics, radon, watershed protection, prohibition of ocean dumping, and large-scale ecosystem health issues such as stratospheric ozone depletion, with EPA being the competent agency in the United States for the Montreal Protocol. In 1989 the Exxon Valdez tanker collision in Alaska and management of the resulting oil spill thrust the spotlight back on the EPA's responsibilities for ecosystem protection.

The 1990s were marked by further progress on a variety of fronts, especially once EPA faced less constraint from a more-sympathetic Clinton administration. The United Nations Earth Summit in 1992 had set a global agenda for sustainability as well as the traditional environmental missions of conservation and pollution prevention. EPA had the power and the mandate to advance the agenda in the United States, with congressional support, from the local level, with restrictions against passive cigarette smoke, to the planetary level, with bans on chlorofluorocarbons to protect the stratospheric ozone layer. New initiatives were seen in remediating "brownfields," industrial sites requiring cleanup for other land uses, which restored contaminated urban sites to productive use and allowed "in-filling" of the urban landscape for better regional planning and management of density.

The public-health mission inherent in the EPA's health-based standards was broadened, expanded, and elevated to an agency-wide priority with the creation of the Office of Children's Health Protection in 1997. Children's environmental health protection became a highly effective policy for crosscutting programs and public education. It was also politically neutral and unassailable, and therefore served for many years as a means of achieving bipartisan and community support for environmental protection initiatives, despite a polarized political environment.

The EPA also launched a variety of initiatives collectively referred to as "right to know," which expanded the information resources available to citizens and advocacy groups and brought transparency to environmental monitoring. As in the case of children's environmental health, this was politically neutral because it made information more accessible to all citizens, whatever their purpose.

During the administration of George W. Bush, EPA was widely perceived as taking a big step backward and becoming less effective. A case in point was the decision not to revise the regulation for permissible levels of arsenic in drinking water, despite new evidence suggesting a higher risk of toxicity than previously believed. Another was the diminishing number of enforcement actions taken by the agency compared to previous years. However, the agency generally maintained forward progress until the last years of the administration, when it was caught in a series of policy maneuvers that were more directly aimed at halting or reversing measures for environmental protection. One was an initiative and proposed legislation that would have amended the Clean Air Act in a bill called the “Clear Skies Act,” which was perceived as weakening air pollution regulation; reducing control of emissions from power plants; and, toward the end, permitting unacceptable emissions and long-range transport of mercury. This initiative stalled in Congress in 2005. The low point for the agency in the 2000s, however, was probably when EPA Administrator Christie Todd Whitman, a former governor of Connecticut, reassured residents about the health risk of inhaled dust during rescue efforts at the site of the World Trade Center after the September 11, 2001, terrorist attacks. Within a short time, she was contradicted by studies suggesting significant respiratory effects. Administrators, of course, are dependent on their staff and expert advice, but in this instance there was also concern over the degree to which the White House may have influenced her remarks at the time, injecting politics into a scientific and health issue.

The environmental laws and regulations administered by EPA are too numerous to describe in this chapter, but some major acts are described in Chapter 12.

The direct role of EPA in occupational health is relatively narrow, but considerable in its sectors. EPA has regulatory authority over the application of pesticides, and as a consequence, its standards and enforcement procedures are critically important in agriculture. (Pesticide manufacturing comes under OSHA.) EPA also establishes

procedures and monitoring programs for workers involved in cleaning up hazardous waste sites.

## **OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION**

The Occupational Safety and Health Act of 1970 (OSH Act) established the Occupational Safety and Health Administration (OSHA) in the United States and its power to set and enforce federal standards for worksite safety and health. The 1970 OSH Act was passed to ensure employees legal protection against occupational injury and disease: “To assure so far as possible every working man and woman in the Nation safe and healthful working conditions and to preserve our human resources.”

OSHA sets occupational-health exposure standards; requires safe work practices and specifically mandates some for serious hazards; requires maintenance of a log of injuries at the worksite (the “OSHA 300” log); requires training and communication regarding hazards; petitions courts on behalf of workers in case of imminent danger; inspects workplaces (although it is required to give advance notice of routine visits); and approves state programs. There are hundreds of standards covering machine guarding, materials handling, toxic chemicals, noise, radiation, exits, walking and working surfaces, sanitation, fire prevention, welding, staging and shoring, scaffolding, operation and design of machinery, and other hazards. If no specific standard is set for a workplace, there is a “General Duty Clause” in the OSH Act that states, “Employers are required to maintain a workplace free from recognized hazards.” Temporary emergency standards can be set until permanent ones are established to protect workers when new hazards are introduced in the workplace.

### **History of the OSH Act and OSHA**

Unlike the Environmental Protection Agency, which was created from the consolidation of several preexisting agencies, OSHA was newly created under its legislation. OSHA is also much smaller and

narrower in its mission, and therefore more vulnerable to political challenges. OSHA is easiest to understand through an appreciation of its history.

Passage of the OSH Act was the result of a long and concerted effort on the part of organized labor. During the 1960s, there was growing concern about dangerous working conditions, and organized labor was at a high point of influence. OSHA was a response to this concern and, in practice, an extension of the oversight relationship already exercised by the federal government over its contractors. The states had been responsible for worker health and safety prior to 1970 by default. The OSH Act allowed individual states to develop their own state OSHA equivalents, as long as enforcement procedures and standards were at least as stringent as those of OSHA. Thus, the adoption of OSHA was more incremental than a break with the past.

The single most important part of the act is probably what is known as the “General Duty Clause,” part of which reads: “[Each employer] shall furnish to each of his employees employment and a place of employment which are free from recognized hazards that are causing or are likely to cause death or serious physical harm to his employees.” In theory, therefore, the act provides a strong mandate for worker protection and includes an implied expectation that hazards will be studied in order to be recognized. However, read another way, the General Duty Clause applies to hazards that cause “death or serious harm” and so may be viewed as intended to apply only to life threatening or potentially disabling risks. At the same time, the original act did not explicitly declare the right of a worker to refuse an assignment that may result in serious injury or death and not be punished for such refusal. That had to come later by regulation, almost as an afterthought, and as a Supreme Court ruling in 1980 (*Whirlpool v. Marshall*). This ambiguity pervades the act and its subsequent history.

Unlike the National Environmental Policy Act, which was passed the year before and created the Environmental Protection Agency, the OSH Act did not create a single, strong agency with enforcement powers, extramural funding capacity, and standards based on

mission-oriented research. It created a group of individually weak organizations with divided responsibilities:

- OSHA, the enforcement agency, was embedded in the Department of Labor, rather than the department responsible for health (at the time, the Department of Health, Education, and Welfare), in order to align it with other labor standards.
- The National Institute of Occupational Safety and Health—a relatively small agency now absorbed into the Centers for Disease Control and Prevention—had a mission of research and training, and was given the mandate to propose new standards, most of which have not been adopted by OSHA.
- The Occupational Health and Safety Review Commission is the appeals body for OSHA citations.

Occupational health and safety standards then in existence had been written by industry associations and professional organizations; they were intended to be voluntary guidelines for management. They became mandatory for federal contractors under the Walsh-Healey Public Contracts Act of 1936, which, among many other provisions, required covered private companies providing services and products to the federal government to adhere to safety and sanitation standards. The OSH Act adopted by reference the Walsh-Healey standards, which consisted of relevant guidelines defining unsafe, unsanitary, and dangerous working conditions, and elevated them to the status of federal regulations for industry as a whole and created an enforcement agency to ensure compliance. By then, many of the standards and guidelines were already dated, however, and coverage of workplace hazards was not complete. For this reason, the initial set of OSHA standards also included then-current voluntary guidelines that represented a “national consensus” in 1970, namely those of the American Conference of Governmental Industrial Hygienists (ACGIH) and the American National Standards Institute (ANSI). The body of health standards

that governs occupational health and safety in the United States is therefore more than forty years old, for the most part, and some provisions may be seventy years old.

Two mechanisms were adopted by OSHA for enforcing chemical safety. One was the Permissible Exposure Level (PEL), a limit on exposure to chemicals that is primarily based on the assumption that level of exposure (such as concentration in air) and duration of exposure can be treated as a product or a cumulative toxicity model in the form of an eight-hour, time-weighted average for the approximately 600 chemicals so regulated. Though approximately true for some chemical hazards, this is not so for many; but the ease of applying this toxicological simplification, the relatively low risk of toxicity at current levels, and the precedent and body of recommended standards of organizations such as the American Conference of Governmental Industrial Hygienists made this approach attractive and fairly robust. (Ceiling levels and Short-Term Exposure Levels—for chemical exposures in which peak exposures are more significant—were supplemental to the PEL approach.) The other approach was the bundled or comprehensive standard, which laid out in one package exposure limits, mandated surveillance, required personal protection, and special issues such as medical removal, as exemplified by standards on asbestos, arsenic, and lead. These three particular standards demonstrate a number of complications because of differences between the general industry standards and the construction standards for asbestos and especially lead; the calculation of allowable exposure for longer work shifts for arsenic and lead; and the medical removal provisions and criteria for return to work for lead, which differ slightly between the general industry and construction standards.

Notwithstanding the long history and general acceptance of the major part of its body of standards, OSHA was immediately ridiculed as being standards obsessed and lampooned in jokes and cartoons, which very effectively undermined its image and mission. Throughout this early period, OSHA emphasized voluntary compliance over stringent legal enforcement, but even so, the organization gained a

reputation for heavy-handed enforcement and had to contend with numerous efforts to legislate exemptions to its jurisdictions. OSHA did exempt employers with ten employees or less from record-keeping requirements as a measure of “regulatory relief” and introduced a consultation service in 1975 designed to help employers, especially in small enterprises, achieve compliance with standards. A brief but apparently baseless scandal erupted during the Nixon administration in 1974 over whether OSHA enforcement was politically manipulated.

By then, the late 1970s, OSHA was getting a rather late start on new standards setting, which was intended to be a major role of the new agency. However, in the latter half of the decade, this became the priority. A watershed event in standards setting was the standard for benzene, which was the subject of a 1980 Supreme Court ruling that required documentation of the significance of a health problem and effectively introduced quantitative risk assessment into occupational health standards setting.

Among other initiatives, OSHA started the New Directions program (now called Susan Harwood grants) in 1978, providing small financial awards to support training employers and workers in occupational health and safety. OSHA also maintains a network of training centers to provide technical education for compliance officers and interested health and safety professionals in the private sector.

Under the Reagan administration, OSHA was a favorite target for allegations of regulatory abuse, but it scored a major victory when the cotton-dust standard was upheld by the Supreme Court, which held that cost-benefit analysis was not an acceptable basis for a health and safety standard. During the first half of the 1980s, OSHA was caught up in the deregulation effort of the time, which diminished its authority again, a process not helped by a series of messy and highly visible regulatory issues that were interpreted as either withdrawal from a commitment to worker protection or professional incompetence.

In order to be seen as responsive to employer’s concerns, OSHA introduced the so-called “regulatory relief” package in the 1980s—a series of policy changes to reduce the perceived burden of federal regulation of industry. Rather than concentrating on eliminating

hazards from the workplace, this policy emphasizes increasing use of protective gear such as respirators and earplugs. OSHA had allowed only temporary reliance on respirators and similar personal protective devices, which is seldom as effective as engineering controls. A more positive development was the Voluntary Protection Program, which enlisted employers as partners in cooperation for occupational health protection and helped OSHA's troubled relations with the employer sector. Outside of OSHA, however, many saw this and later partnership programs, such as the Alliance Program initiated during the George W. Bush administration, as effort recognition programs that did little to improve enforcement where it was actually lacking. In 2004 a report by the General Accounting Office questioned the effectiveness of such programs.

In 1991 the agency promulgated the Bloodborne Pathogen Standard in response to rising rates of occupational infection from hepatitis B and the risk of HIV among health care workers. This move was unique in having been directed by Congress in a rider to legislation. The standard appears to have been one of its most successful, paving the way for universal precautions and later amendments for the prevention of needlestick injuries (and the Needlestick Safety Act of 2000) and precipitating a major and sustained decline in transmission.

During the Clinton administration, OSHA's emphasis shifted to assistance in compliance with regulatory standards, building on the consultation program that provided technical assistance to employers who asked for it. Skepticism on the part of employers and reluctance to have attention drawn to their issues, even with the intent to avoid enforcement activities, reduced participation in this program.

In 2000, after years of preparation, OSHA faced its most severe test when it promulgated the standard on ergonomics, a well-documented regulation for the prevention of musculoskeletal disorders. (The ergonomics standard did combine low-back pain and repetitive strain injuries, which have different characteristics, into one regulation, which in retrospect may have been a mistake.) It was ferociously opposed by employer groups led by the United States Chamber of

Commerce, and despite a favorable review of the underlying science by the National Academy of Sciences, it was ultimately repealed by Congress in what was viewed as a stinging rebuke.

OSHA remained a weakened agency for the duration of the Bush administration. Its regulatory initiatives came primarily in the form of voluntary compliance and cooperative programs for voluntary adherence to guidelines, which largely vitiated the role of the agency in enforcing mandatory compliance with regulations. An example of this weakness was the manner in which OSHA, which had the responsibility to ensure the safety of workers so as not to compound the tragedy, was marginalized in the immediate aftermath of the September 11, 2001, terrorist attacks on the World Trade Center in New York, where its role was largely restricted to handing out respiratory protection on-site for rescuers who could be persuaded to use it. Later, it played almost no role in the investigation of the assault using anthrax, although almost all of the victims were exposed in the course of their work duties.

By the close of the Bush administration, OSHA seemed to be struggling mightily with two seemingly straightforward issues. The first was the appropriate hazard standard to be promulgated for prevention of a newly recognized serious lung disease (colloquially called “popcorn lung”) caused by diacetyl and seen in commercial popcorn operations. The second was enforcement of regulations dealing with the imminent hazard presented by combustible dusts, specifically with respect to explosion hazards in the sugar and grain industries, for which there has been an applicable standard since 1987. Although enforcement of a major action arising from a fatal explosion in a sugar factory resulted in a large and highly publicized fine to the employer, what seemed to be an obvious and intolerable hazard appeared to be mired in confusion as OSHA discovered that seventeen different standards could apply. The issue was met with a National Special Emphasis Program rather than the emergency standard that was recommended by the Chemical Safety Board.

Under the Obama administration, new as of the time of writing (2009), OSHA is likely to have much more support in both Congress and the executive branch. OSHA has many more options open to it

than in the past, among them the development of generic and class standards, adoption of recommended exposure limits (RELs) proposed by NIOSH, a return to process safety management as a key regulatory tool, a closer alignment with public-health agencies, the introduction of innovations and simplified approaches such as control banding, harmonization of occupational health standards with the European Union (where the process has continued to advance), and greater engagement with training and education.

OSHA has made several efforts to deal with the inertia of regulation by adopting generic or comprehensive standards for categories of exposure (one unsuccessful effort would have provided a template for regulation of carcinogens) and for policies that broadly support regulation—most notably the 1983 Hazard Communication Standard (also called “Right to Know”), which had the effect of raising worker awareness and capacity for self-protection.

OSHA also initiated an awards program for exemplary employers called the Safety and Health Achievement Recognition Program (SHARP), which involves a voluntary inspection.

## **Enforcement**

Standards are enforced through inspections. Once a job hazard is encountered, it should be brought to the attention of the employer. If no action is taken, an employee may contact the OSHA area office, where the employee will be able to file a complaint in writing. All complaints are evaluated, and priority is assigned by the perceived severity of the hazard. On the basis of this, OSHA staff members decide whether to conduct an inspection. OSHA policy has been to inspect all potential serious violations within three working days.

The first priority for inspections is imminent hazard, when there is an immediate risk and action must be taken quickly to prevent the potential for death or serious injury. When there is evidence for this, the compliance officer normally requests voluntary abatement by the employer, and if it is refused, the officer seeks a court injunction against the employer to suspend operations until an inspection can be

completed. The second priority is the investigation of fatalities or catastrophic events resulting in three or more employees hospitalized. These must be reported to OSHA by the employer within eight hours. The third priority is to respond to formal complaints from employees or other interested parties. Programmed or routine inspections come last and are prioritized by the safety record of the employer and the industry, or by the presence of potential hazards of concern.

Inspections for enforcement purposes are surprise visits; employers are not notified in advance unless there is a fatality reported, a clear reason why the inspection would require it, or an imminent hazard that should be corrected before an inspection is possible. There is an initial meeting with management (a workers' representative may be present) called the "opening conference," followed by meetings with responsible managers and employees, and a visit to the location of the suspected violation. Following the inspection, there is another meeting with management called the "closing conference," at which preliminary findings are discussed. OSHA then issues a report in the form of citations, which are outlined in a letter citing the specific violations and a date by which the hazard must be abated or the violation corrected. A fine may also be assessed. There may be follow-up inspections to determine whether a hazard has been abated.

The employer has the right to dispute the citations but must do so within fifteen days. The first level of appeal for citations by OSHA compliance officers is through the OSHA area director to the Occupational Health and Safety Review Commission, which uses a system of administrative judges with a final level of review by the commission itself. The counterpart appeals system for state OSHA plans is generally similar but involves state review boards.

Should a violation be found, fines are determined by how serious a risk the violation presents to workers. There are categories of violation:

- "Other-than-serious" violations
- Serious violations

- Willful violations
- Repeated violations
- “Failure to abate” a prior violation

Citations on grounds of imminent danger are very rare; when they are made, the shop is closed down immediately until the danger is corrected. Serious violations are those that have substantial probability of causing death or physical harm. If management shows “good faith” in trying to keep things safe, there is a reduction in the penalty and a further reduction for small businesses. Additional fines can be added for various reasons, such as failure to post notice of the violation within the required time period (three days) until it is abated. Additional fines and possible criminal sanctions are levied for falsifying or concealing records and reports, or for assaulting or interfering with a compliance officer.

Various factors are weighed before citations and fines are issued, such as the size of an employer’s business, the seriousness of the violation, the apparent good faith of the employer, and any record of prior violations. After all these factors are taken into consideration, penalties are frequently minimal. Because it often costs considerably more to correct a hazard than to pay OSHA fines, abatement dates may be missed or ignored by employers. An employer may be assessed a civil penalty of not more than \$1,000 for each day during which a violation continues past an abatement date. Extensions beyond the abatement date are common. During this time, employees may continue to be exposed to the hazard.

On the other hand, there is no limit on the dollar amount of a fine. In 1987 federal OSHA imposed a fine of \$1.6 million on Chrysler Corporation’s Newark, Delaware, plant for multiple violations. In 2005 the biggest fine in OSHA history, \$2.78 million, was levied on Cintas, a laundry company in Tulsa where a worker was dragged to his death in a dryer and multiple willful violations were discovered. This is very unusual, however, and such cases are widely interpreted as efforts by OSHA to show visible strength.

Under Section 11(c) of the OSH Act, an employer is prohibited from firing or discriminating against any employee “because of the exercise by such employee on behalf of himself or others of any right afforded by this act.” Specifically included in this section are the rights to file a safety and health complaint, to institute a proceeding under the OSH Act, and to testify at any hearings. Despite Section 11(c), however, many workers fear that their employer will nonetheless find grounds to discipline or even fire them on other grounds or on pretexts if they exercise their rights. Unfortunately, this appears to happen often enough to confirm such fears. Once an employee is singled out as a “troublemaker” by management, it is easy to accumulate real or imagined complaints that may lead to termination, ostensibly on other grounds.

### **State OSHA Programs**

The enabling legislation for federal OSHA assumed that most states would eventually assume partial responsibility for occupational health and safety under broad federal supervision. Most were slow to do so. At the time of writing (2009), there are now twenty-six state programs, of which three cover only public employers. State regulations, as permitted under the OSH Act, are to be at least as strict as the federal law and may be even stricter in setting standards and enforcement policies. Much depends on financial resources available, however.

The story of the state agency in California is illustrative. California had been considered at the forefront of worker protection, largely due to the progressive administration in the state department. In 1987 the California state OSHA (CalOSHA) was rescinded by the governor from the budget and dissolved, although the state contribution to the agency represented only an \$8 million expenditure per year. Fifty percent of CalOSHA’s budget came from the federal government, and fifty percent came from the State Department of Industrial Relations. As budget cuts on the federal level reduced the federal contribution, maintenance of the state agency had become

increasingly more difficult. After the governor's decision, California was covered by an office of federal OSHA that included many former CalOSHA employees. In an initiative referendum in 1988, voters restored CalOSHA, but by then many of the staff had changed careers or left their jobs at what was briefly the California office of federal OSHA in search of stabler careers. The net effect was a considerable weakening of the agency that became CalOSHA again in 1991. CalOSHA was reestablished and restaffed and went on to regain its clout and to become an assertive regulatory agency again. However, its close call with oblivion underscores just how fragile state OSHA programs may become.

### **The Legacy of OSHA at Forty Years**

How one views OSHA depends on who one sees as the victim in the enforcement of occupational health protection. Worker advocates see workers as the victims, with companies inflicting suffering on them through unhealthful working conditions and time pressures. This does not necessarily mean that workers are solidly in support of OSHA as the advocate of their best interests, however. Workers do not always see OSHA or other issues from a "worker" point of view. They hold their own political opinions and often view it as an example of regulatory excess, depending on their views. Employers and managers, on the other hand, show more consistency in their opinions against "government regulations." Their advocates typically see the employer as the victim, with OSHA, the workers, unions, and government in general inflicting excessive and unnecessary regulation, which causes delays, frustration, and increased costs. Intrusion into the production process is not easily accepted by the business manager or owner who expects to control that process.

OSHA has been widely seen as too weak on enforcement. What OSHA has accomplished generally, however, is the ever-greater and more-sophisticated appreciation of the problems of workplace health and safety. Nothing has been as important as the message that has come from the existence of OSHA: change the workplace, not the worker.

Until the passage of the OSH Act, personnel or industrial relations managers rarely had more than the most rudimentary safety and health training. Few industrial relations and personnel courses or texts contained more than a mention of theories of accident causation. Hazard control was delegated to maintenance or engineering departments. Top management was little involved in hazard controls. Supervisors in these departments were directly responsible to line management, whose major concern was production and not the improvement of working conditions. The safety director rarely had much influence or support from top management unless a major accident occurred. Many companies regarded accidents from a cost standpoint and retitled their safety and health director as the "loss control" supervisor. The old approach tended to consider safety issues only after an accident or illness had occurred. Then, attention was often focused on laying blame, especially on the worker who was seen as carelessly committing an unsafe act, rather than on correcting for human limitations in performance in an intrinsically unsafe environment. It may be said that attitudes toward safety on the part of management lag at least a generation behind the knowledge base of safety science (see Chapter 6), and by much more in some sectors. (Shipyards and hospitals are particularly slow in this regard.)

Today the situation is far from ideal, but it is much improved, as it should be after forty years. A vigorous effort has been made to clean up the workplace. A keystone of this effort has been the direct training of workers to recognize hazards, to participate in their control, to have the means to control them, to monitor hazards in the workplace, and to know and to exercise rights under the law.

## **MINE SAFETY AND HEALTH ADMINISTRATION**

The Mine Safety and Health Administration (MSHA) is an agency within the Department of Labor. MSHA separated from OSHA in 1978 and took with it responsibility for occupational health and safety regulation in mines, including underground and surface mines

and quarries. MSHA administers the Federal Mine Safety and Health Act of 1977. However, it had a much longer antecedent history prior to union with OSHA in 1970, and for most of its statutory life, it was part of the Bureau of Mines of the Department of the Interior, a status that ended with the OSH Act of 1970.

MSHA inspects underground mines four times a year and surface mines twice a year to assess occupational hazards. MSHA also has responsibility for approving respirators, the principal personal protective equipment for respiratory hazards.

In 1978 the research functions attached to MSHA's predecessor agencies were moved over to the National Institute for Occupational Safety and Health (NIOSH). In addition to supporting the certification of personal respiratory protection, which is the one MSHA responsibility that affects general industry, NIOSH conducts surveys for coal workers' pneumoconiosis through a medical surveillance program to assess progress in eliminating pneumoconioses.

Mining has always been among the most dangerous industries in the United States, as it is elsewhere. The history of coal mining in particular is inextricably entwined with occupational health and safety as well as society's understanding of occupational health risks over several generations (see Chapter 25). Mining presents many specialized hazards—mostly associated with underground mining—which was the major reason MSHA was removed from OSHA and restored as a separate agency. These specialized problems include dust exposure; tunneling and sinking shafts, and preventing their collapse; roof support and preventing roof collapse, roof falls (rocks detaching and falling from overhead), and outbreaks (rocks separating from the sides); pockets of explosive gas (methane, sometimes with hydrogen sulfide); explosive dust; diesel exhaust; safety and reentry during intentional explosions; ventilation; lighting; electrical hazards; fire (and generation of carbon monoxide); flooding; and mine rescue, which is a highly specialized field of emergency management. Underground mines are, of course, confined spaces and present many problems of access and egress; it is easy to get trapped during a mine accident. Surface

mines, on the other hand, are usually more like large construction sites in their hazard profile.

Dust exposure is a particular and characteristic hazard, including silica in both underground coal and “hard rock” (noncoal, further divided into “metallic” and “nonmetallic” in the industry) mines that require excavation and drilling through rock, and coal dust in coal mines; pneumoconioses, especially silicosis and coal workers’ pneumoconiosis, are therefore the signature occupational diseases in the mining industry. Historically, the high priority of respiratory hazards in mining has led to MSHA, rather than OSHA, becoming the leading agency in certifying respiratory protection equipment, although NIOSH conducts research on standards and testing for respiratory personal protection equipment.

The federal effort to regulate safety in mines began in 1891, with an act that set minimum ventilation requirements and addressed child labor, establishing a minimum age of twelve. The largest fatal industrial incident in U.S. history occurred in a coal mine in Monongah, West Virginia, in December 1907, when 392 men and boys working in the Fairmont Coal Company died in an underground explosion. Later in the month, another mine explosion in Dary, Pennsylvania, killed 239 workers. It was followed two years later by a mine fire in Cherry, Illinois, that killed another 300 miners. These mining disasters killed mostly immigrants from Italy and Hungary, not native-born Americans, and so had less impact on public opinion than they otherwise might have had, but the magnitude and frequency of the disasters still motivated Congress to pass legislation establishing the Bureau of Mines in the Department of the Interior, which was the predecessor to MSHA.

The Bureau had the responsibility of encouraging the exploitation of mineral resources and supporting the critically important extraction industries, as well as regulating safety. Its safety authority was limited to research and education. However, during the 1920s and 1930s mining fatalities and injuries increased, demonstrating that this weak approach was not working. The Bureau did not acquire the power to inspect mines until 1941, at which time a code of regulations was also adopted by Congress, largely through the influence of the United Mine Workers

and its president, John L. Lewis. The Federal Coal Mine Safety Act of 1952 strengthened the agency in the coal sector, where occupational injuries were particularly common, severe, and often affected more than one worker in one incident; but the act did not give the Bureau the power to levy fines for noncompliance with regulations, leaving enforcement possible only through civil litigation. In 1966 these provisions, which in the 1952 act applied only to certain coal mines, were extended to all coal mines and to non-coal mines as well, but with still-weak enforcement measures. Meanwhile, sporadic mine disasters continued to plague the industry and mining communities, where the burden of health effects was also being felt in the form of coal workers' pneumoconiosis (colloquially called "black lung" and other respiratory diseases). In 1968 another mine explosion, this time in Farmington, West Virginia, very close to the Fairmont site, killed seventy-eight miners and abruptly drew attention to the extreme risk of coal mining. In 1969, the year before the OSH Act, the Coal Mine Safety and Health Act of 1969 was passed partly in response to the Farmington incident; the act required regular inspection of surface and underground coal mines based on risk, and it increased enforcement powers, including the authority to levy fines and to lay criminal charges.

In 1973 the Department of the Interior created the Mine Enforcement and Safety Administration (MESA) in order to separate occupational health and safety regulation in mines from the Bureau's larger mandate to support the extractive mineral industries. The Federal Mine Safety and Health Act of 1977 amended the 1969 coal act; brought together regulation of coal and non-coal mines; shifted MESA into the Department of Labor, as a sister agency to OSHA; and transformed it into MSHA, with greater enforcement powers. The act also created a Federal Mine Safety and Health Review Commission, parallel to the Occupational Health and Safety Review Commission, for appeals of MSHA citations and actions.

The 1980s and 1990s were characterized by a number of contentious enforcement cases, some of which involved allegations of fraud in monitoring dust levels and most of which reinforced the traditional stance of mine owners and operators as a sector among

the most resistant to regulation in American industry. Mining is driven by commodities prices, which are highly volatile. The revenue generated by a mine, coal or non-coal, peaks when the price is high, but this generally does not last long. (The early 2000s, with the expansion of industry and the rapid development of China, have been an exception, and in the current economic climate as of this writing, prices have plummeted.) Mine owners are motivated to take profits in the short term while they can and otherwise to keep costs at a minimum in order to keep the operation sustainable when commodity prices are low. This economic motivation is often put forward as a reason for the resistance of mine owners and operators to changes in work practices and regulation.

During the George W. Bush administration, MSHA, like OSHA, emphasized partnership and voluntary programs with employers. Over the 2000s (as documented in a story in the *New York Times*, March 2, 2006), MSHA levied generally fewer citations for much smaller amounts and did not collect a large fraction of these fines.

During MSHA's three decades, it has generally been considered disappointing in its performance. Whether from resource constraints, lack of political will, or a few highly publicized incidents of irregularities, MSHA has been widely perceived as excessively deferential to the industry and weak in enforcement. Since the creation of MSHA, fatalities and serious mine injuries have declined overall, but beginning in 2005, another series of disasters began—Sago (WV), Alma (WV), Darby (KY), Crandall Canyon (UT)—reversing the trend and raising questions regarding the adequacy of regulation and whether MSHA had acted on prior indications of deficiencies in mine safety. In 2007 NIOSH reported that the incidence of coal workers' pneumoconiosis, which had been falling, was rising again, in younger coal miners.

## CONSUMER PRODUCT SAFETY COMMISSION

The Consumer Product Safety Commission (CPSC) is responsible for protecting the consumer, who may be in the role of a worker or a private citizen, from hazards associated with consumer products,

including but not limited to durable goods (such as appliances), building materials, mobile homes, tools, household (but not industrial) chemicals, and toys, which have been of particular interest and concern in its history. CPSC does not regulate food or medicine, medical devices, or cosmetics (Food and Drug Administration); pesticides (Environmental Protection Agency); automobiles and other vehicles (National Highway Traffic Safety Administration, Department of Transportation); boats (U.S. Coast Guard); alcohol, tobacco, or firearms (the latter three are regulated by the Bureau of Alcohol, Tobacco, and Firearms, which for historical reasons of collecting taxes is part of the Department of the Treasury).

CPSC's main activities include the following functions:

- Issuing and enforcing mandatory standards
- Developing voluntary standards with industry representatives
- Ordering the recall of products identified as hazardous
- Ordering the repair of products identified as hazardous
- Consumer education
- Market surveillance to identify hazardous products in commerce
- Surveillance of imported products at ports of entry
- Surveillance of injuries related to consumer products through a reporting system based in 100 hospital emergency rooms (the National Electronic Injury Surveillance System)
- Surveillance of deaths related to consumer products
- Liaison with foreign governments to harmonize and raise safety standards (especially in manufacturing and distribution processes)

CPSC can replace a voluntary standard with a mandatory standard only after determining that the voluntary standard is inadequate or is unlikely to be complied with. CPSC's most visible actions are product recalls. These may occur on an expedited basis

with cooperation of the manufacturer or distributor through the “Fast Track” program, or more conventionally through legal action.

CPSC was created in 1972 when Congress passed the Consumer Product Safety Act, which directed the new agency to “protect the public against unreasonable risks of injury associated with consumer products.” This act was amended by the Consumer Product Safety Improvement Act of 2008. CPSC also administers six other federal laws:

- Children’s Gasoline Burn Prevention Act of 2008 (portable gasoline containers)
- Federal Hazardous Substances Act of 1960 as amended several times (household and consumer products)
- Poison Prevention Packaging Act of 1970 (child-resistant packaging)
- Pool and Spa Safety Act of 2007 (preventing suction-caused entrapment and drowning of children in draining pools)
- Refrigerator Safety Act of 1956 (requiring that the door can be opened from the inside to prevent confinement and suffocation)

The Federal Hazardous Substances Act has been a particularly important source of regulatory authority and has been used to ban carbon tetrachloride, potentially dangerous aerial fireworks, strong liquid drain cleaners, and products containing cyanide and asbestos. It also mandates labeling for consumer products containing hazardous chemicals.

Although a freestanding federal agency, it is relatively small compared to other regulatory agencies (420 employees to manage 15,000 categories of consumer products). Its administration is divided into functional units, of which the operational (as opposed to support) offices are Compliance and Field Operations, Hazard Identification and Reduction, and International Programs and Intergovernmental Affairs.

CPSC has established itself as a leading innovator among federal agencies, creating a single portal (“one-stop shop”) for managing recalls by all federal agencies with the authority to order them, an Early Warning System to identify and disseminate information on new products and emerging safety issues, a National Injury Information Clearinghouse, and electronic systems to transmit information on product recalls to media and participating Web sites electronically. The agency’s generally proactive approach has served it particularly well in recent years during a period of challenges involving the risk of imported toys contaminated with lead, use during natural disasters of portable generators (which when used indoors create a serious risk of carbon monoxide toxicity), and toy magnets (which, when more than one is swallowed, can attract one another and pinch the bowel, causing necrosis and bowel infarction or intussusception).

## **OTHER U.S. FEDERAL AGENCIES**

In the United States, workers in sectors that are outside the mainstream of general industry or defined by special hazards are often regulated by agencies other than OSHA. This is primarily to ensure access to relevant technical expertise but also, as in the case of aviation and the military, to ensure that there is not a conflict between worker protection and carrying out sensitive duties.

The Department of Transportation (DOT) requires regular fitness-to-work evaluations for drivers, pilots, and seamen. Also, branches of DOT require drug testing for these transportation workers and pipeline operatives. This is discussed in Chapter 12.

Within DOT, the Federal Aviation Administration (FAA) has regulatory authority over the occupational health and safety of aircraft-maintenance and flight personnel. The FAA is responsible for occupational health protection of flight personnel in civil aviation (mostly pilots and flight attendants) during periods of operation, which are defined strictly as when the first member of the crew arrives and the last departs, whether the airplane is in flight; on the ground, either in motion or holding; or whether engines are running

or off. At other times, and at all times for the maintenance and ground crews, occupational health is the responsibility of OSHA. FAA has promulgated regulations regarding health and safety in airplane cockpits, including lighting, seat-belt use, ventilation, and smoke evacuation, but not noise, which has been controversial. FAA regulations are most visible in terms of evaluating pilots' fitness for duty, discussed in Chapter 18.

The Department of Defense (DOD) has a very large responsibility for occupational health and safety, covering the protection of active-duty personnel (including those with combat status) as well as civilian employees. The military maintains numerous facilities for the maintenance, development, and production of weaponry, support equipment, and vehicles. A military base may function more as a location for industrial-scale technical activity than for deployment and training. The same exposures common in industry generally are common in the military, and many military activities also involve potential exposures that are unusual or exotic, such as torpedo fuel (nitrate compounds that cause vasodilation and hypotension if absorbed), particles from jet aircraft brakes (beryllium alloys), and classified agents of warfare. Each branch of service maintains occupational health and safety units and supervises workplace safety in facilities under its direct control. In the past, DOD resisted being placed under the authority of other federal agencies and did not accept OSHA standards as binding. This impasse was resolved in the late 1970s by executive order, and in the 1980s DOD developed a strong intramural occupational health and safety program consistent with OSHA regulations.

The nuclear industry is another sector that has been carved out for special management of occupational health protection, a legacy of the Atomic Energy Act of 1954. The Department of Energy (DOE) is the federal agency responsible for occupational health and safety related to the nuclear industry's radiation hazards and for DOE national laboratories and test facilities. The Nuclear Regulatory Commission (NRC), which licenses nuclear plants and nuclear medicine in health care, is responsible for establishing exposure

standards for ionizing radiation, chemical hazards that arise from radioactive materials (for example, uranium, which is nephrotoxic), and plant conditions that may affect radiation safety. The NRC also sets exposure standards for the diagnostic and therapeutic use of radionuclides in nuclear medicine. OSHA has authority over other occupational health hazards, such as chemical exposures. In practice, the agencies take a team approach. The Department of Transportation manages regulations on transport of radioactive materials; EPA regulates ambient exposure in the environment; and the Food and Drug Administration manages food irradiation and pharmaceuticals. This is a complicated area, and specific guidance should be sought if an issue involving radiation falls within the responsibility of an OEM physician.

The Department of Agriculture has programs to promote farm safety and health. It also regulates laboratory-animal housing and management. Because laboratory-animal handlers have a high rate of occupational injuries and allergies, this is a significant responsibility with implications for worker health protection.

## **CANADIAN REGULATORY FRAMEWORK**

### **Environmental Health**

Under the Canadian constitution (devolved from the British North America Act), environmental regulation and conservation is a provincial responsibility, as is health. The role of the federal government is therefore circumscribed.

The lead federal department (ministry) for environmental issues is Environment Canada. Federal responsibility is primary in issues involving inshore or offshore fisheries (a responsibility shared with the Department of Fisheries and Oceans), aboriginal lands, waters, national parks, and a few issues that are defined as being of “national concern,” such as management of toxic substances, migration of contaminants through the ecosystem, protection of endangered species, management of Canada’s engagement in international

treaties and conventions, and so forth. Federal action on the environment otherwise is generally undertaken in partnership or collectively with provincial and territorial agencies, which vary in their capacity but are generally consistent in environmental regulation. Environment Canada also has responsibility for research, monitoring, and prediction of weather, climate, and environmental trends, making it Canada's lead agency in matters of climate change.

Provinces (of which there are ten) and territories (of which there are three) have ministries of environmental protection, which are responsible for environmental standards, monitoring emissions, local habitat and ecosystem conservation, environmental assessment, inspections, and enforcement. OEM physicians who practice in Canada or who have responsibility for Canadian operations are advised to find out about local regulations and environmental regulation in the relevant province or territory.

Canada operates primarily by a system of guidelines that represent targets for attainment rather than mandated standards, although mandated standards may also be adopted by the provinces. Guidelines for standards setting are negotiated by a federal-provincial working group and are implemented as emissions standards at the provincial level. These include the Canadian Drinking Water Guidelines and similar guidelines for recreational water use and air quality. Remediation standards for contaminated sites (including hazardous-waste disposal sites and brownfields) are governed by a set of consensus guidelines negotiated directly by provincial ministers responsible for environmental protection: the Canadian Council of Ministers of the Environment.

## **Occupational Health and Safety**

Government services in occupational health vary widely across Canada because those services are primarily a provincial (or territorial) responsibility. Each of the ten provinces and three territorial governments as well as the federal government (primarily on behalf of its own employees) has its own occupational health and safety agency, freestanding in Quebec but attached to departments of labor

in other jurisdictions, for a total of fourteen jurisdictions. In British Columbia, uniquely, occupational health and safety is joined with the Workers' Compensation Board. These agencies are responsible for administering the individual occupational health and safety acts relevant to their jurisdictions, conducting inspections and investigations, surveilling occupational injuries and illnesses, recommending standards, and supporting training or research. Most acts cover only nonagricultural workers. They vary somewhat in services, coverage, and enforcement policy. There is some uniform legislation for some occupational health and safety functions, primarily the hazard communication and "right-to-know" through the Workplace Hazardous Materials Information System, which covers labeling, worker education, and access to material-safety data sheets.

The federal government is responsible for occupational health and safety for federal employees and employees of Crown corporations and employees in industries that operate across provincial boundaries (including banks, transportation, telecommunications, and grain elevators), which together cover 10 percent of the Canadian workforce. The federal government is also involved when a role exists for it in facilitating voluntary and cooperative efforts. For example, the federal government sponsors the Canadian Centre for Occupational Health and Safety, a Crown corporation (government-owned, nonprofit company) that primarily disseminates information on occupational hazards but has no legal authority in issues of standards-setting enforcement or national policy.

Separate from the agencies (except in British Columbia) are the workers' compensation boards, which are Crown corporations that serve the dual functions of claims processors and insurance carriers (rather than allowing multiple private insurance companies). In general, there is little interaction between the occupational health agencies and the workers' compensation boards, with the latter focused on claims, rehabilitation services, and insurance services, except for exchanging statistical information. In a few cases, efforts have been made to innovate, such as Alberta's "window of opportunity program," which provided a financial incentive for

employers in selected industries that improved their performance in the short term.

The OEM physician with responsibility for operations in Canada should become familiar with the provincial and federal legislation and regulations affecting the industry.

## **General**

In Canada, reimbursement for health care services to physicians is a responsibility of provincial governments through their provincial health-insurance plans, with federal contribution from transfer payments and regulation. The system functions as a government-administered health insurance company paying physicians on a fee-for-service basis. (Hospitals are funded through direct annual allocations.) Remuneration for occupational physicians for services not directly related to the treatment of patients must be from third parties.

Government-based provincial health plans, whether the provincial health plans or workers' compensation, do not usually reimburse physicians for work done related to occupational health protection or occupational health services related to work capacity, such as pre-placement or periodic health surveillance programs. These costs are the responsibility of the employers. In Ontario, government, labor, and industry have supported a network of occupational medicine clinics: the Occupational Health Clinics for Ontario Workers. In Quebec, some occupational health services are provided by departments of the "community health centres," a network of service facilities supported by the government and usually attached to regional hospitals.

## **MULTILATERAL AGENCIES**

Several important agencies in North America represent shared responsibilities and governance over environmental issues. These agencies do not have direct regulatory authority, but they are influential in actions taken by the national regulatory agencies.

Because they are official bodies and quasi-regulatory in their functions, they are included in this chapter.

The International Joint Commission (IJC) is a binational organization governed by appointed commissioners from the United States and Canada. Since the Boundary Waters Treaty of 1909 was ratified by the parties, the IJC has had the responsibility of overseeing and proposing nonbinding recommendations for environmental issues that straddle the U.S.-Canada border. Historically, most of its work has been directed at preventing pollution, conserving resources, or mitigating degradation of watershed for bodies of water, such as the St. Lawrence River and the Great Lakes. There are standing bodies, or “boards,” for the major rivers and lakes. The Great Lakes Water Agreement of 1978 has dominated the agenda of the IJC in recent years, and the organization has been very active in monitoring water quality, studying the interaction of air deposition and bodies of water, and relating the ecosystem to human health. Although primarily concerned with water issues, one of the IJC’s first important health issues involved air (involving a smelter in Trail, British Columbia, in 1910) and much of its most recent activities include monitoring and advising on transboundary issues involving air pollution, particularly through its International Air Quality Advisory Board. The IJC can only take up issues that are referred to it by both national governments and has only advisory powers; but its recommendations have been influential, and it has had an active program of community education and has supported research and education relevant to ecosystem and human health, including sponsoring a task force of health professionals.

The U.S.-Mexico Border Health Commission is an advisory body created in 2000 by international agreement; it consists of state health officers (four American and six Mexican states) or their delegates and appointed representatives. It is concerned with environmental health issues within 100 kilometers on either side of the international border. To date, it has dealt with tuberculosis, lead in traditional pottery, and air pollution from a power plant, among other issues. It acts mostly as a forum and through advocacy, coordination, setting achievable

objectives, and public education. The commission has close ties with the El Paso regional office of the Pan American Health Organization and the Association of State and Territorial Health Officers, the organization representing state health officers in the United States.

The Commission for Environmental Cooperation of North America (CEC) is a trilateral organization established in 1993 by the North American Agreement on Environmental Cooperation, a side agreement of the North American Free Trade Act. The CEC monitors environmental issues in the three member countries (the United States, Canada, and Mexico), particularly with respect to trade and the impact of international trade on resources, environmental quality, and potential conflicts. It receives and investigates submissions from the public on issues but has no enforcement or regulatory authority.

## **RESOURCES**

Information from and on regulatory agencies is abundant on the Web and in both professional and nontechnical publications. However, much information and opinion tends to be ephemeral, issue specific, and often partisan. Regulations also change and agencies evolve in response to politics and social demands. Readers are advised to do their own Internet searches on any significant issue to ensure currency and accuracy, and not to rely on books for anything other than historical guidance.

Bassett WEH, ed. *Clay's Handbook of Environmental Health*. London: Taylor & Francis; 2002.

Collin RW. *The Environmental Protection Agency: Cleaning Up America's Act*. Westport, CT: Greenwood; 2005.

Fletcher SR, Copeland C, Luther L, McCarthy JE, Reisch M, eds. *Environmental Laws: Summaries of Major Statutes Administered by the Environmental Protection Agency*. Hauppauge, NY: Nova; 2008.

Silverstein M. Getting home safe and sound: the Occupational Health and Safety Administration at 38. *Am J Public Health*. 2008;98(3):416–423.

## **NOTEWORTHY READINGS**

Bingham E. On the need for reform of the OSH Act. *Am J Ind Med*. 1992;21: 891–893.

Cheit RE. *Setting Safety Standards: Regulation in the Public and Private Sectors*. Berkeley: University of California Press; 1990.

- Doniger DD. *The Law and Policy of Toxic Substances Control: A Case Study of Vinyl Chloride*. Baltimore: The Johns Hopkins University Press / Resources for the Future; 1978.
- Lofgren DJ. *Dangerous Premises: An Insider's View of OSH Enforcement*. Ithaca, NY: ILR Press / Cornell University Press; 1989.
- McCaffrey DP. *OSHA and the Politics of Health Regulation*. New York: Plenum Press; 1982.
- Smith RS. *The Occupational Health and Safety Act: Its Goals and Its Achievements*. Washington, DC: American Enterprise Institute for Public Policy Research; 1976.

*This page intentionally left blank*

# **25 SOCIETY, CRITICAL SCIENCE, AND ETHICS**

Occupational and environmental medicine (OEM) is unique in medicine in many respects, among them its dual emphasis on the individual patient, worker, or community resident and the population, workforce, or community. Because of this concern for protecting the health of people both as individuals and in communities, occupational medicine is rightly called “public health for the employed population,” and environmental medicine is rightly considered to stand at the interface between medicine and public health.

OEM is also unique in that it is driven by changes in demography, the economy, and technology more than medical science. Occupational medicine practice scarcely changes with the introduction of a new antibiotic or anti-inflammatory agent, but it changes profoundly as the workforce ages, as employment patterns change, and when a new chemical hazard is introduced into the workplace. Occupational medicine is characterized by an innate conservatism scientifically because the biomedical science on which it relies must be validated and well accepted; its epidemiologically derived knowledge base requires rigor; and its procedures must be legally and ethically defensible after experience in practice. Environmental medicine is more responsive than occupational medicine to advances in biomedical

science because of the rapid advances of bioindicator technology, molecular epidemiology (a term that is falling from favor), and toxicogenomics, which examines gene expression and gene-toxicant interactions.

Coming from their different histories and traditions, the merging medical fields of occupational medicine and environmental medicine are harmonizing their different relationships to the community and different orientations toward social issues. As the fields continue to knit together, their common alignment will come from the notion of OEM as a medical field or specialty that bridges to public health and serves defined populations, whether in the form of a workforce or a broader community. For the time being, however, the two poles of OEM are best discussed separately, after a brief introduction to the concept of “critical science,” which applies to both sides of the field.

### **OEM AS “CRITICAL SCIENCE”**

Another concept of OEM is that it is one of several scientific or medical fields that document and seek to correct problems of technology and society.

The economic success of technology and applied science has had mixed social blessings. OEM, like the environmental sciences and risk assessment, is at a deep level a critique of the technology and society of the day. Such a critique is not in automatic opposition to progress. Rather, it is a means of correcting the errors of technology and of economic growth before they threaten to impede progress and to undermine human values. To see how this works—and how it has come to define a new form, or “mode,” of science—requires an understanding of why science and technology are not the same thing, and a realization that science can legitimately study and critique the effects of technology and the economy.

Basic science and applied science are the two commonly accepted “modes” of scientific research, the latter being associated with technology. Scientists and academics talk about the difference between basic science and applied science, but the public rarely follows the

distinction, considering science and technology to be two sides of the same coin. To decision makers, it often sounds like the distinction is merely a justification for grant funding for basic research. However, the differences between the two are important outside of academic debate because they have different functions altogether.

Basic science pieces together a worldview in part by documenting the unexpected, which expands society's vision beyond narrow applications and documents the unpredictable. This is the mode of science that fills in the uncharted white spaces of the metaphorical map, the mode on which society falls back when confronted with the unexpected and from which come so many fortuitous observations and unanticipated applications, as well as an understanding of mechanisms. It is basic science that pieces together a world picture. Basic science cannot have a ready answer for every question as it arises, but such answers may come closer when investigators have been allowed to extend their investigations into areas that appeared improbable for application before the question arose.

Applied science and problem-oriented technology, on the other hand, are mission oriented and involve problem solving. They create value and innovation. However, they are by their nature poorly equipped to anticipate problems because they are goal oriented and therefore one-sidedly searching for the predictable.

The emergence of a third mode was the result of a new way of thinking. The idea emerged in the 1970s that a systematic evaluation, using the tools of basic science, of the negative risks of technology could create a new mode, which was named "critical science" by its principal theoretician, Dr. Jerry Ravetz, of the University of Leeds (United Kingdom). He described it as a new approach that critically evaluates and analyzes the impact of technological developments within the context of the real world (not the engineer's world of simplifying assumptions), and which draws primarily on basic science and collaborative critical research to predict, document, and correct physical and human problems of technological origin.

The concept of a "critical science" describes much of the role that OEM at its best plays in society. For its more inquisitive practitioners,

OEM is very much in the mode of a critical science. It addresses issues of occupational health and of environmental health risk as they arise from changes in technology and the economy. The objective is to identify and then to correct these problems for the good of humanity and the sustainability of society. The whole point of a critical science is to identify and correct social problems, particularly those that arise from technology. As such, it is profoundly in support of progress and change, because it seeks to remove obstacles to progress, to prevent adverse and unintended consequences, and to correct what has gone wrong.

Occupational medicine, at one pole of OEM, is a critical science because it seeks to identify, document, and correct the problems of the workplace, which arise from technology, work organization, and economic development. Environmental medicine, at the other pole, is a critical science because it seeks to identify, document, and correct the environmental problems caused by human activity, which reflect changing technology (which introduces new hazards while also correcting others), failure to use technology (for example, to provide clean water in the developing world), and economic change. Environmental medicine is also closely aligned with ecology and environmental sciences, the disciplines Ravetz first cited as examples of critical science. Like its sister critical sciences, OEM, as a whole or at its two poles, uses the methods of basic science and research to investigate practical problems, and it depends on basic science to identify previously unsuspected or undocumented problems. Applied science and technology build on the insights and mechanistic studies of basic science, but they do not do very well at anticipating unintended consequences. Disciplines such as epidemiology and toxicology identify and evaluate problems, and that is, of course, the first necessary step in their correction.

The notion of OEM as critical science is not unlike Rudolph Virchow's admonition that:

Medicine is a social science, and politics is nothing else but medicine on a large scale. Medicine, as a social science, as the

science of human beings, has the obligation to point out problems and to attempt their theoretical solution: the politician, the practical anthropologist, must find the means for their actual solution.

This idea, which neatly embraces OEM, applies equally well to public health. At its best, OEM is in the same tradition with Virchow's social medicine in treating health as a human right—a means of achieving full expression in one's personal life and social role—and an end result of the struggle for fairness, freedom, and material security. It cannot solve every problem, of course. It is embedded in a social, economic, and management system that has a responsibility to be responsive on a larger scale and to change when necessary.

Both poles of OEM will be discussed separately—environmental medicine first because it fits more obviously within Ravetz's definition.

### **Environmental Medicine as a Critical Science**

Health professionals in North America are often deeply concerned with the environment in their personal lives, and some physicians have been prominent as activists. The environmental movement in North America often speaks in terms of health, and health is a driving concern for the public in shaping the response to environmental problems. However, the term "environmental health" has a different meaning in the health professions, including public health and OEM, than in the environmental movement. Health professionals perceive "environmental health" as pertaining to environmental determinants of human health. To environmentalists, and to many members of the public, "environmental health" is as likely to mean "the health of the environment." Some have dubbed the latter "ecosystem health." The public generally sees human and ecosystem health as mixed together and both fundamentally related to social justice and fairness.

Public health professionals and environmentalists often find themselves frustrated in communication with one another. The concept of environmental determinants of health as expressed by the public

health professional may appear very limiting and even self-serving to an environmentalist accustomed to the big picture of “the environment,” or of ecosystem stability. On the other hand, the concept of ecosystem health or health in a holistic sense, as used by environmentalists, may appear to be vague, indirect, and unprofessional to a health professional accustomed to thinking of specific exposures, calculable risks, and specific disorders.

This difference in understanding becomes most contentious when health issues are presented as evidence of environmental degradation on the basis of evidence that may not meet the rigorous standards of epidemiology. If one takes the environmental movement, especially its more radical elements, literally on a technical level as an alternative interpretation of exposure to toxic and physical hazards, it sometimes appears to bend the science and, at the extreme, to be antiscientific and even irrational. The discrepancy between the approach of scientists, environmental health professionals, and OEM physicians, on the one hand, and environmental advocates and activists on the other is not unlike the difference in standards of certainty described in Chapter 23 for scientific and for medicolegal issues. Environmentalists are using a standard often described as the “precautionary principle,” which for this purpose can be described as requiring action to protect against risk if there is any evidence of an environmental hazard—even if there is scientific doubt about the evidence—because the consequences are considered to be unacceptable. (There are other definitions of the precautionary principle.) This scenario is played out over and over again in issues such as the risk of cancer from common herbicides used on lawns and parks, the risks of electromagnetic fields, and popular assumptions regarding the risk of getting certain diseases, usually those with unknown causes, such as multiple sclerosis, autism, and sarcoidosis. At its best, the precautionary principle leads to continuous improvement, such as green technology, without the need to justify every step along the way. In public health and OEM, however, it often leads to priorities that are questionable in public health practice—such as emphasizing the putative risk resulting from disinfection byproducts in water over

control of waterborne infectious-disease risk—and wrong public health decisions—such as discouraging breastfeeding because of the putative risk resulting from PCB residues when breast milk is known to have unequivocal benefits to infants.

There is also a tendency among naive environmentalists to use health concerns as a rhetorical device to raise awareness of a skeptical public to an issue that is considered to be critical to the environmental agenda. Environmentalists may perceive themselves to be sounding the alarm on important health issues that illustrate the link between the environment and human health, and therefore called upon to motivate people to political or social action. To health professionals, however, attributing health problems to environmental causes without rigorous evidence sounds alarmist, and the science often seems suspect. It is also threatening to the hard-won and fragile credibility that public health professionals and OEM physicians always seek to protect.

For all these reasons, health professionals are often inclined to disassociate themselves from environmentalists, at least as professionals, for fear of losing public trust. Environmentalists, who usually are primarily concerned with getting action on what they consider to be the big issues, may therefore see their health professional colleagues as hair-splitting, timid careerists who shy away from the root causes of health problems, which in their view lie in the perverse relationship between human society and the environment.

What is the most appropriate role for physicians in environmental issues? Obviously, well-trained and focused physician investigators will continue to make substantial scientific contributions to the field. A relatively small number of highly motivated physician activists will also continue to make contributions in public awareness and on specific issues. An individual physician may feel an ethical obligation to speak out on environmental issues, but for most there is a limit to their comfort level. Physicians are generally most comfortable playing a personal role in environmental issues when there is a practical objective to be achieved that is related to a health problem they know, when there is a clear health connection supported by evidence,

and when a clear distinction is made between the professional role of clinician and the personal role of citizen. They generally show a preference for incremental rather than radical change and a preference for advocacy over activism.

Physicians' awareness of new hazards and the process of disease as well as the influence that physicians have from the profession's high public acceptance and esteem give physicians a special role. That role suggests several opportunities for action, depending on the physician's training and commitment:

- Work within organizations, such as companies and government, to correct and prevent environmental problems
- Document through research the health effects of environmental degradation
- Endorse, support, and advocate for an effort to improve human health by moving toward environmental sustainability
- Work within the health sector to ensure that it is environmentally responsible in its operations and facilities
- Act responsibly by example, both as an individual practitioner and employer
- Work with organizations outside the health sector to help them understand issues of the environment and health
- Work within medical organizations locally, nationally, and on an international level to help them understand issues of the environment and health and to take responsible policy positions

Physicians need to take account of their standing and of the respect that they receive from local communities, and to use this respect responsibly to act as advocates for ecosystem health and human health. The first obligation of the physician who becomes active in environmental issues must be to prepare him- or herself—as thoroughly as possible for a new medical practice—to learn enough about environmental sciences and health issues to speak responsibly and knowledgeably.

On the other hand, organized medicine may also speak collectively for the individual physician. What are the appropriate roles of the profession as a whole and of “mainstream” medical organizations?

Assuming that physicians are confident that the organization demonstrates sound judgment, it is reasonable to present a collective voice for physicians that is cautious, authoritative, and concerned but also effective. Certain organizations, such as the International Society of Doctors for the Environment, the Canadian Association of Physicians for the Environment, and the International Physicians for Prevention of Nuclear War and their affiliates (Physicians for Social Responsibility in the United States, and Physicians for Global Survival in Canada), will play a role in the vanguard of medical participation in these issues.

### **Occupational Medicine as a Critical Science**

Just as environmental medicine fits Ravetz’s description of a critical science for environmental issues, the other pole of OEM, occupational medicine, is dedicated to preventing, identifying, and correcting problems and unintended consequences in the workplace, as well as to improving conditions of working life. This extends to using the workplace and the employment relationship as a platform to enhance the health of workers and to contribute to the community. Occupational medicine is a critical science for working life, and is not limited to hazards on the job. However, it is in the realm of workplace hazards that the function of occupational medicine as a critical science is most obvious.

For any form of occupational health protection to be effective, the employer and the worker need to know what is going on. That is the crucial role of occupational medicine as a critical science. A fundamental issue in occupational health in general is who controls the workplace and whether employers have an obligation to share governance of issues such as health and safety with the workers they employ. In the United States, it is the employer who controls the workplace and the working environment and who must therefore

bear full responsibility for occupational health protection. This creates a tension between the employer's prerogatives and ownership and the right of workers to be protected or to be given the means to protect themselves. This tension is irresolvable in a country—the United States—that defines itself in terms of entrepreneurship, private ownership, and creative destruction. However, the more informed, educated, trained, and involved workers are, the more likely OHS is to be effectively monitored and managed. That is why worker participation features so prominently in the ethics of occupational medicine (discussed below). But without knowledge about workplace hazards and information from the particular workplace, neither employer nor worker can take effective action. This is the role of occupational medicine as a critical science.

### *The Imperfect Emerging Economy*

New hazards will continue to emerge as technology develops and exposure patterns change in the workplace. Europe, Japan, North America, and some parts of the middle-income, developing world have moved into a largely “postindustrial” society. Services in this new economy are as highly valued and important as manufactured products. Among the features of this postindustrial society is an emphasis on information as a commodity and automation of much manufacturing, which comes to resemble agriculture in that it produces at high volume but with a small workforce. Much production is outsourced overseas, currently mainly to China, and then, by China as well as North America, to other countries with lower wage levels. There is a potential for new or rediscovered occupational hazards with each new industry, whether it be high-technology electronics; advanced materials such as composites, ceramics, or biotechnology; or the shift work and stress that accompany a global trading network, requiring split-second decision-making and devastating consequences for error.

The old hazards that result in predictable injury and disease, with their burden of disability and insecurity, do not disappear, however. They go overseas to affect workers in other countries where the

work is outsourced. They persist in low-margin industries that are difficult to outsource, such as construction. They hide in the informal sector or in industries with a history of abusive practices, often disproportionately harming immigrants and other vulnerable populations. They are perpetuated by ignorance or lack of awareness in small enterprises that lack the means to improve conditions or do not see the point. As the economy changes, it gets harder to find occupational hazards, but not because they have disappeared. They become less visible and affect workers who are least able to protect themselves and earn the lowest wages.

Workers who are vulnerable and are not provided with the occupational health protection they need are called “underserved workers.” Underserved workers are most often thought of as well-defined groups, such as recent immigrants, disadvantaged minorities, illiterate workers, and the economic underclass. However, it may be more useful to think of underserved workers as an underclass workforce defined by broad social issues. These issues include the concentration of dirty and dangerous work by class, geography, and sector as well as ethnicity, migration status, and income level, into poorly documented, less accessible corners of society largely hidden from the majority population and the media, often in marginal industries where employment is unstable.

Occupational health problems are relatively invisible already because they are spread through many industries and many workplaces. Occupational health has not coalesced into a social issue that people get excited about. Occupational health issues are perceived, wrongly, by the public as the concern of a special interest, such as organized labor, or a narrow problem in a particular industry or employer. The deaths or illnesses tend to occur one at a time, or in small clusters. They tend to occur quietly and are obscured by the bureaucracy and statistical categorization of agencies like workers’ compensation boards and occupational health and safety regulatory agencies, necessary as those statistics may be.

It is a fundamental tenet of occupational medicine that old hazards always reappear in new technologies. For example, in the manufacture

of semiconductors and other electronic components, gallium arsenide is a very important constituent. There were cases of arsenic intoxication among semiconductor workers when the industry began, and a great deal of denial and disbelief at first that such a “clean” industry could cause this problem. However, arsenic is a very old hazard, well known and well studied, and its risks and measures to control exposure were understood long before the semiconductor industry was invented. However, the people who pioneered the manufacture of electronic components were not thinking in terms of the mining and smelting industry. For them, this was a bright, new, unique technology starting from the beginning. Truly, old hazards never quite go away; they just reappear in a new technology.

North American society seems to have entered a twilight zone in occupational health and safety. The European Union has taken the lead in occupational health protection. The United States and Canada are passing from an era in which there were certain clear ideas of what the major issues were—and these issues had strong evidence of commitment from government and public policy—into a new and ill-defined place, an uncharted territory through which both countries are still feeling their way.

The problems of greatest concern have changed over the last two decades, although it is difficult to be sure how much of the change is a real difference in changing impact on society and how much is a change in perception. Asthma and reactive airways, solvent-induced neurological problems, repetitive-strain injuries, soft-tissue injuries, passive smoking, injury control, and violence emerged as the growth areas of interest. Soft-tissue injuries, particularly in the context of repetitive-strain injury, have been primarily perceived as an issue in ergonomics rather than medicine. The old, characteristic diseases of occupational medicine, such as silicosis and noise-induced hearing loss, seem not to inspire any great interest anymore because of the (false) assumption that they have disappeared.

For historical reasons, so much of the American social insurance system is tied to employment (health insurance, pension benefits, disability, unemployment insurance) that managing these issues far

outweighs managing work-related hazards as a practical problem, from the employer's point of view. General health issues, such as health insurance, disability management, wellness, and managed care, now command much more attention from employers and government than occupational health and workers' compensation. There is little interest and less reward in going looking for occupational health problems in obscure places or unpopular corners of the economy. In the new economy, occupational safety could fall to an even lower priority because of the perception, incorrect as it is, that the injury problem is solved.

Occupational medicine is the medical arm of occupational health, which in turn is part of the broad field of occupational health and safety. The mission and fortunes of OEM are therefore linked to the priority given to occupational health and to the attention paid by management and by society to occupational health and safety (OHS) in general. In recent years, just before the time of this writing (2009), OHS went through a long period of marginalization and neglect, especially on the federal level. This was justified at the time on the questionable grounds that injury rates were falling and that industry was overregulated. The former was hard to be sure of, given changes in the way that injuries were counted during the period, and the latter was certainly arguable. In such a policy and management environment, it was hard to sustain interest and momentum for health protection at the level of individual employers, especially for small enterprises. OEM physicians had to adapt in many ways, but they mostly maintained their positions and continued to do well, if not flourish.

Hazard control interventions are intended to keep something bad from happening, and it is always more difficult to document that bad events have not happened than it is to show that positive changes have occurred. Control of workplace hazards is difficult to sustain as an enthusiastic interest of management, but once the controls are in place they are difficult to change. Safety measures, to prevent incidents of trauma, are usually easier for management to comprehend and to evaluate than health protection. Safety, although it can be very

sophisticated as a field, tends to be less technically complicated in practice because the measures required to control safety hazards are relatively straightforward by comparison to health protection. Safety science is more a problem of behavioral engineering than technical engineering. At the heart of sound safety practice is the difficult problem of motivating and sustaining behavior.

OEM physicians must be aware of trends in OHS in general, and their implications for the field and for practice. Interest in health, as opposed to injury, runs in cycles, just as does interest in occupational medicine. There is a seesaw in management interest between safety and health. At intervals, the seesaw tilts rather sharply toward safety, usually following highly publicized incidents, especially those involving fatalities. Concern over cancer, hazardous chemicals, and other, more subtle effects usually tips the balance back the other way, but does so more slowly, bringing health back as a major management concern over time. It is important for the OEM physician to realize this dynamic, because it is often reflected in management perceptions, proposed reorganizations, and cooperation between safety and health professionals. Safety departments in major companies are often more or less autonomous and generally resist being incorporated in the same administrative unit with medicine and occupational health. The emphasis seesaws between safety and health; however, it is ultimately pointless for there to be a pendulum swinging between safety and health. Protecting workers is a common goal and a shared professional commitment between the two fields.

A positive change in the workplace has been that OHS is increasingly considered standard operating procedure. This is critical: if health and safety are not built into line responsibility, they simply will not be a priority. It is not appropriate to make the health or safety professional “responsible” for health protection: it has to be a commitment of the organization followed through with policies, procedures, and work practice. For this to work, in turn, occupational health has to be part of performance evaluation. Supervisors need to know that they can be fired for failure to perform diligently in health and safety issues. Today, performance review policies generally are

more accurately characterized as likely to deny a promotion to a supervisor who fails miserably in occupational health and safety, but such a failure hardly ever leads to a supervisor's termination unless a worker is killed or seriously injured.

Occupational health and safety professionals in industry are always afraid of losing resources. In the 1980s and 1990s, many corporate medical directors in large companies resisted the introduction of health promotion programs, because they were concerned that hazard control would get short shrift. Others adopted health promotion, incorporated it into their departments, and were able to keep their focus during the ensuing popularity and then plateau of enthusiasm for worksite health promotion.

Periodically, management and social fads distort priorities in health protection and affect the mission of occupational health and safety, including OEM. These come and go, and like all management fads, they capture the imagination of managers who are looking for a new, simple solution. An illustration of the general principle was the movement called "quality of working life" (QWL), or "conditions of work," which built up a substantial following in the 1970s. QWL emphasized the conditions of work and personal control over the working environment. It took a very general approach, assuming that the major determinants of a worker's sense of well-being have to do with the immediate social and physical environment of the workplace and the amount of control the worker has over personal comfort and work pace. QWL advocates were frequently shortsighted in assuming that the principal occupational health hazards had been controlled already. Some QWL advocates, who by and large were isolated from the realities of work outside of offices, even went so far as to say that occupational health was no longer pertinent to the modern world, because industry was no longer as hazardous as it was during the Industrial Revolution. In effect, QWL advocates were substituting for substantive hazard control what they believed to be the nonspecific and probably nonexistent "Hawthorne effect" (see Chapter 13) of enhanced worker productivity and satisfaction stimulated by management attention. QWL, if done well, may have

provided the worker with greater control over the working environment. It is more likely that it served little purpose other than the manipulation of the working environment to achieve management objectives. There was also a considerable commercial potential in selling management on aesthetic changes that improve office and plant decor. Both worksite health promotion and quality of working life were aggressively marketed, but occupational health was not.

At the time, QWL was seen as reflecting good and enlightened management. On the other hand, occupational health implies a need to solve problems that perhaps should not be present in the first place. No supervisor or manager wants to admit that the workplace for which they are responsible is hazardous. No supervisor or manager wants to believe that a worker was injured due to a preventable hazard in the workplace that they overlooked or allowed to exist. It is much easier for managers to deny the problem, to turn to a quick fix, to provide incentives for underreporting injuries, or to assign responsibility to someone else without giving them real authority. It takes courage and professional commitment to accept that a workplace is hazardous and to take steps for occupational health protection. That is why occupational health services that work well and that have the backing of employers are uncommon and to be commended. It is also why the management tools that focus attention on occupational health and safety—such as written policies, worker and management training, performance evaluation of supervisors that includes safety performance, and audits—are critical to sustaining an OHS effort overall and maintaining support for the OEM physician.

Occupational health is often at a disadvantage for management attention and priority for allocation of money and resources. There is a very real danger that when the heat is off, occupational health services will be rolled back and replaced by the more popular and less expensive health promotion and QWL measures. Control of workplace hazards is difficult to sustain as an enthusiastic interest of management, but once the controls are in place it is possible to show management the ongoing benefits. Hazard control requires resources and sustained commitment; it requires constant vigilance; and it is

usually not perceived as exciting or responsive to the workers' psychic needs in the way that both health promotion and quality of working life are. In the end, however, it is the heart and soul of occupational health and safety and of OEM.

## **HISTORICAL PERSPECTIVE**

The history of occupational health illustrates the importance of OEM and the historical contributions of science in a critical role, documenting the problems of technology.

Western society moved through several distinct phases along the technological path to modernity. Europe and its North American colonies started out as agrarian societies, but England and Scotland soon developed urban centers that depended on technology rather than trade alone. Britain was first to enter a very difficult, technologically driven Industrial Revolution. Some parts of the world are only now entering theirs and are having similar difficulties, although history never repeats itself exactly.

North Americans, having prospered and become as much a fountain of innovation as was Britain (and France) in the past, now live in a very different technological age than the one that began with the Industrial Revolution. North American society has moved into a phase beyond the Industrial Revolution in which it is trying to solve the lingering abuses of an earlier technological age and at the same time to move on to another, qualitatively different and more egalitarian era of more sophisticated technology and organizational management. A natural concomitant of this progress has been an aversion to risk. Each generation became more affluent and acquired the technological means and the purchasing power to insulate itself, or at least a majority of the population, from risks to a greater and greater degree—first the risk of starvation, then poverty and privation, then discomfort and inconvenience, and now risks that would have been tolerable or even acceptable to this generation's parents. Societies become more risk averse as they grow more affluent, age in their demographics, and mature socially. North American society is now in

the enviable position of being able to worry about small risks of cancer and about reproductive hazards, which would have been inconceivable to this generation's grandparents.

However, the aversion to risk is selective. It focuses on hazards that are visible (usually through the media), associated with dread outcomes, and affect the middle class, but it largely ignores the persistent hazards of the workplace and problems such as contaminated water, which is still the major environmental hazard worldwide (see Chapter 7). The social aversion to risk was not enough to motivate measures to prevent global climate change, the greatest environmental risk of all, despite years of warnings. The gains of the last century may not be held. The economy is again restructuring and is, at the time of this writing, in deep trouble worldwide. It has been said that the now-passing generation of "baby boomers" may be the first in North America to have expectations of a life less affluent than their parents.

Until the present generation, Western society has tended to equate the progress of technology—with its promise of more and more engineering and organizational breakthroughs—with progress in terms of better lifestyle and better individual opportunity economically. (In another era, this was the ideology of the party known as the "Whigs.") It would be a mistake to think that progress is uniform, however, and naive to think that technological progress necessarily brings a better life for all. The early Industrial Revolution caused enormous dislocation in the fabric of British society at the time. It emptied the English and Scottish countryside of people and impoverished half the United Kingdom. It caused severe urban impaction, to a degree never seen before, and caused epidemics and plagues of injuries and illness of historic proportions, exceeding even the contemporary developing world, where the situation is bad enough. This led to a profound cynicism regarding technology's role in society, perhaps best expressed in the 1936 silent movie *Modern Times* starring Charlie Chaplin. This film focused on what technology seemed to be doing to the individual during the Great Depression, which represented a failure of progress. The world had created technology, it

seemed, and now that technology had not only failed but was creating its own negative effect on the world.

### **Coal Mining as a Mirror**

There is no industry in which these issues are clearer than coal mining, not because it was necessarily typical but because it showed all of these problems so vividly and was so closely linked to technological change as the dominant energy source. The fabric of society in coal-mining communities was very thin and very much dominated by the hard work of mining coal. The social effects of living in a one-company town were limiting. When the pay is good (or better than the alternatives) and family is rooted to the place, workers accept the risk because they have to. The miner who went down into the mines in the 1880s had little protection in the way of safety: there were no hardhats, and light was available only from an oil-fired lantern that gave off a very dim light (and caused “miner’s nystagmus,” a now-forgotten occupational disease). The work was hard and hazardous. It was not unusual in coal-mining communities of the late nineteenth century and early twentieth century for a woman to be widowed three or four times in her lifetime, one husband after another dying in mine-related accidents.

It was standard practice in those days for canaries to be brought down in the mines by miners as an early detection device for toxic gas (carbon monoxide or hydrogen sulfide). If the canary dropped dead off its perch, the foreman knew that the mine was unsafe and the miners would evacuate it. Today, the term “canaries” is used in occupational health to indicate those workers who become ill first, because they are often the first indication of a problem in that particular workplace. The canary became an important metaphor in the English language and one of the few metaphors drawn from occupational health that is regularly heard in common speech.

Technological progress sometimes results in occupational disease. Technological progress does not necessarily mean progress in the health and working environments. For example, the prevalence of “black lung,” or “coal-worker pneumoconiosis,” and “silicosis” in

coal miners increases the higher the respirable dust level. In the 1880s, holes were drilled in coal seams using fairly primitive equipment. Less dust was generated, and those miners tended not to get “black lung” as often as those in later years. Of course, mining was very primitive and dangerous anyway, because of other hazards. The introduction of more effective equipment brought ever more serious safety hazards, including unprotected open gears and belt drives, which required people to be sitting next to operating machinery. A drilling machine, substituting for the hand drills of the early eighteenth century, operated at a much higher speed. The dust that was generated was much greater in concentration and much finer in size than the dust the manual drill could have produced. The effect of this was “black lung.” Today, there are many practical methods to keep dust levels down, and the drill bit is kept wet.

Child labor was also extremely common in those days. A child about seven or eight years old might have lived underground for most of his life. He might get up in the morning and go down to the depths of the mine around 5:30 a.m. and spend his entire day doing various jobs or pulling small carts removing coal from the mine. By the time he finished for the day and went topside, he may have gone without seeing daylight for days on end. Another job performed by young boys in England and America was to pick pieces of slate out of the coal as it came by on a conveyor belt. For girls, the same story was repeated in the textile mills. As children grew older they could do heavier jobs, move away from home, and live in dormitories. This was done for ten hours a day with a half-day on Saturday. They went to church on Sunday. The life of adults was not much better in those days.

Mining today is vastly different than in those days, and although it is much safer than it was, it is still dangerous. Underground mining still takes its place among the most hazardous occupations, as evidenced in modern China and other developing countries. Early on, however, progress brought its own problems, and improving technology introduced new and different hazards. Hazards without effective controls place workers at risk. If the controls are not used they

may as well not have been invented. But if hazards are identified and controlled by means of effective protective measures, both the employer and the worker benefit from technological progress and society moves forward. The environmental costs of coal mining—from water pollution to mountaintop removal—also serve as a demonstration that occupational and environmental health are linked, often in indirect ways.

The history of coal mining is ancient and may seem passé today, but there are lessons in it for all time.

## ATTITUDES TOWARD OCCUPATIONAL MEDICINE

OEM and OHS, so rich in science, history, and social relevance, has never truly broken through to receive appropriate recognition as a discipline and remains marginalized, although much progress has been made (see Chapter 1). This problem is also shared by public health and preventive medicine in general. One might think that occupational health would escape the trap because of its demonstrable usefulness and well-documented cost-effectiveness. The conventional explanation is that OEM has failed to make its case and that those in the field are ineffective at communicating its value. However, there is another explanation: occupational health has been marginalized not because it has failed but because people want it to fail and consider it, at some level, immoral!

The explanation is based on arguments made by George Lakoff in his 1996 book *Moral Politics: What Conservatives Know That Liberals Don't*, which discusses assumptions and cognitive frameworks in which political concepts are judged. Lakoff's contribution is to observe that attitudes toward private matters and public issues alike are conditioned by two belief systems in North American society, which he calls the "Strict Father" and the "Nurturing Parent." Lakoff suggests that this reality is more complicated than labels of "liberal" and "conservative," representing two opposed, highly structured, deep, and contradictory sets of attitudes modeled on experience with the family. The two views of the world may be held on different

issues by the same person and different views may be applied to different issues.

The Strict Father emphasizes character above all and considers it to be an attribute, rather than a behavior, with separate and evaluable components of virtue, discipline, resourcefulness, responsibility, and enterprise. The Strict Father values focus and clarity in goals. The worthy person must demonstrate that self-restraint, self-discipline, and self-reliance come from within. The Strict Father has a duty to protect only when children are immature, but once the children are grown, he has a duty to be strict with them. “Character building” is a responsibility of the Strict Father. Character building is a process that fosters self-discipline and self-control. Rewards that are not earned are immoral. To deny a child the opportunity to take risks and to suffer the consequences is immoral. It denies the child the opportunity to grow and become self-reliant. Value is placed in single-mindedly achieving an objective, one goal at a time, because this demonstrates perspicacity and character. The Strict Father paradoxically values, at the same time and to the same degree, both heritage (and the validation that comes from precedent) and the freedom to take risks and to innovate. The defining freedom for the Strict Father is not the freedom to move in any direction of one’s choosing but the freedom to take a risk that will result in personal advancement. Risk, in this paradigm, is evidence of character, in that one puts one-self on the line for progress and personal gain. The Strict Father bases his moral framework on a model of family life and morality that is reflected in the nuclear family and in child rearing. It places the burden of responsibility squarely on the individual, in return for which the individual has liberal (small *l*, in the classical sense of individualism) freedoms and opportunities without many obligations to others, except for the primary responsibility, which is to allow his or her children (or dependents) the same opportunity to build character. The Strict Father also places heavy emphasis on tradition, loyalty, and preservation of the social order.

The Nurturing Parent treats personality and achievement as learned and mutable in response to society, not as inherent or forged by

uniquely personal experience. Good character will arise in an environment of social responsibility, human rights, diversity, dignity, health and safety, and freedom of expression and action. Interdependence is the reality in this view, and recognition of interdependence defines character. The expression of that good character is empathy for others, a desire to help others, communitarian action, and the healthy recognition that in order to take care of others, it is important to take care of oneself. Self-reliance and the nurturing of others are inseparable in a healthy personality. The Nurturing Parent considers complete independence from others to be an illusion. Protection is a form of caring, and it is reasonable for people to protect the weakest among them when the hazards are beyond one's control. Although personal advancement may be a good thing, it should not take place at the expense of the common good, nor should any member of society be so arrogant as to believe that one's own efforts in isolation, without the social framework, could have accomplished anything worthwhile.

It may seem that there could be no more divergent systems and that between them they cover all possible political and social ground, but really these paradigms are two extremes of one dimension and far from exhausting all possibilities. The two paradigms actually have much in common. Both are based on the idea that the family is a proper model for society and that the education of a child is the proper model for social relationships. The goal of this developmental model in both cases is an independent, self-directed adult capable of playing a constructive role in the community. They do not describe the relationship of adults to one another very well, because they are based on child-rearing models. Both paradigms abhor neglect of responsibility and assume that behavior reflects moral values. Both place great emphasis on the *meaning* of behavior, interpreting actions, beliefs, and rhetoric as consistent or inconsistent with their value systems, whatever the issue and however great the stakes riding on the issue. Both paradigms are profoundly concerned with character as an inherent quality that determines how a person will act under all conditions, which does not allow the idea that behavior in context reflects a specific set of conditions.

Lakoff uses his powerful analysis to demonstrate how these paradigms can explain almost every contradiction in, specifically, American political life. The implications are obvious for certain key issues where moral or “wedge” issues are involved, particularly those with a moral dimension, such as abortion rights and gun control.

Lakoff extends his analysis to perceptions of work and the value of labor. He observes that under the Strict Father paradigm, employment is not an equal relationship. The work relationship reflects the moral order: the employer is in a legitimate authority position, and work is associated with risk and personal advancement and is subject to the same meaning with respect to character and self-discipline as any other aspect of life. Work is a service to the employer, and pay is the reward for obedience in providing this service as instructed, at a fluctuating price set by the market in relation to the demand for the worker’s skill. Workers are therefore expected to be obedient and to welcome the opportunity to prove themselves, demonstrate character, and advance themselves. A high, even brutal, level of competition is welcome and builds character. On the other hand, the Nurturing Parent considers that work should be a social relationship of equals, in which labor is exchanged for fair compensation based on the value added by the labor, and that the workplace should support individual aspirations and provide opportunities for growth in a secure environment.

There is much of value to occupational health in Lakoff’s analysis. It explains why attitudes toward occupational health and safety, and toward individual injured workers, often reflect a seemingly irrational hostility and resistance on one extreme and empty rhetoric on the other.

Occupational health protection is difficult to fit into the paradigm of the Strict Father. Hazards are inevitable to the Strict Father, and to face them squarely and to take risks is a sign of character. If the risk does not pay off and there is an injury, it is unfortunate, but such is life. The injured worker should receive treatment, fair indemnification by a strict standard that does not foster dependence, and then be prepared to move on, accepting that the market value of his or her labor has diminished. The privilege of returning to work must be

earned by effort and loyalty. Indemnification for permanent disability is suspect and destructive to the moral order because it is a reward unearned by effort.

The Nurturing Parent, on the other hand, sees risk taking in the workplace as corruption. In this view, risk is assumed by the worker primarily for the benefit of the employer. Hazard, which is always preventable, is imposed on the worker, and for the worker to accept unsafe conditions is a sign of servility; the right to refuse dangerous work is therefore fundamental to a free society. If an injury occurs and it is serious, it is a tragedy that may affect many lives—those of the worker and his or her family—and the employer should make the worker whole through indemnification, admitting responsibility, and demonstrating concern. Temporary, partial, or total disability is a loss of opportunity to live one's life and diminishes the injured worker's ability to function fully in society. Permanent impairment is a state of need, deserving of compassion. People in need should be helped for the duration of their need. The injured worker or the disabled applicant has a right to equal access to work, after a fair test of ability to do the job with accommodation if necessary.

From these widely divergent points of view come differences of opinion on what to do in the workplace. Safety measures beyond the most obvious may be seen by the Strict Father as evasion of self-discipline on the part of the worker and, on the other hand, as the responsibility of the employer by the Nurturing Parent. Regulation, in the form of occupational health standards, may be seen by the Strict Father as interference with an enterprise that is an engine for social betterment and an expression of moral behavior; the same regulation may be seen by the Nurturing Parent as a necessary but entirely insufficient bureaucratic evil that treats symptoms rather than causes. Enforcement of occupational health standards may be seen by the Strict Father as harassment and an immoral compromise in the opportunity to take risks, whereas the Nurturing Parent may see such enforcement as a failed exercise that is never adequate, equitable, or comprehensive. Government-sponsored programs that support employers' assumption of "internal responsibility" or "partnerships"

(in which employers police themselves, such as OSHA's Alliance program described in Chapter 24) may be seen by the Strict Father as government interference in the workplace, but at least as an admission that employers should be allowed to govern themselves; such programs may be seen by the Nurturing Parent as an untrustworthy, immoral collusion between government and industry in an effort to avoid enforcement. To the Strict Father, regulatory agencies are stupid, but to the Nurturing Parent they are corrupt because they do not go far enough. Regulatory policies either promote moral hazard by removing the need for self-discipline, as seen by the Strict Father, or they are tainted by compromise, as seen by the Nurturing Parent.

These attitudes intensify when workers' compensation is considered. Workers' compensation is generally a minor cost of doing business, but claims often provoke ferocious resistance on the part of some employers, often out of proportion to their exposure to financial loss, and this resistance is met by outrage on the part of workers' advocates. To a Strict Father, workers' compensation benefits are a nonproductive social expenditure that undermines discipline and the social order. To the Nurturing Parent, such benefits are an insufficient social support system and a means by which employers evade responsibility. Workers' compensation satisfies neither.

In Lakoff's analysis, issues related to occupational health and workers' compensation, almost uniquely, generate negative interpretations on both sides. Neither side likes occupational health and both consider the field marginal at best and immoral at worst.

Is there a way forward? Perhaps there is, but only if the paradigms are challenged.

Both paradigms are predicated on the same basic idea: that the work relationship reflects the nuclear family. This is nonsense. Workers are not children and employers are not parents. Neither the workforce, the firm, nor the nation is a family. Occupational health is a matter between (mostly) adults with shared interests in health, economic productivity, and security. It can be dealt with apart from these metaphors of family, with an emphasis on prevention, benefit, and human rights. Occupational health and safety—and by extension

OEM—is a normal part of working life, neither a favor to servants nor a trophy for organized labor. It is a branch of public health and a part of daily life, ideally carried into the community in a seamless regard for safety and health.

## **ETHICS IN OCCUPATIONAL MEDICINE**

Ethics is the normative, or prescriptive, study of behavior that arises from social interaction. It is a way of guiding behavior by deriving a set of rules that govern duty and appropriate action that would be considered good or right. When an ethical principle is violated, the reaction is expected to be proportionate to the magnitude of the infraction, and it is accepted that in a diverse society ethical principles may vary and may require periodic re-evaluation.

Morality, on the other hand, is the law handed down by a higher authority, in religious tradition a deity. In secular society, Immanuel Kant proposed that the whole of humanity, acting in its own collective interest, serves to create morality instead of a deity. Moral laws are not negotiable, because they come as an indivisible whole. The moral system must be accepted or rejected in its entirety. The system is received from an absolute authority, so the elements of the system are presumed to be immutable. The moral system may rest on an initial premise, such as the existence of God, such that its elements all flow from this assumption or belief, and it may be structured to be internally consistent. Such moral systems are fragile, resting on deductive logic, and their parts are inseparable. Sometimes moral systems conflict with social needs and freedoms, but that does not reduce their power. In fact, a tension between social freedoms and a moral system may reinforce the moral system, as believers discipline themselves in the exercise of what they believe to be their imperiled faith. When morality is violated, the response is disproportionate, because violation of one moral law threatens the integrity of the structure as a whole and so cannot be permitted.

Laws are human inventions that are part of a code of behavior more concrete than ethics, incorporating both natural laws of inadmissible

behavior and rules for the settlement of conflicts that could, in a different context or society, be resolved differently by honest arbitrators (see Chapter 23). Ethics is, in a sense, the test of the goodness of a law. The law is one means to determine what society as a whole believes is appropriate behavior under certain conditions but it is also full of compromises, balances, and adjustments to the exigencies of the “real world.” This section only provides general guidance on issues of ethics and touches briefly on law. Whenever legal issues are or might be involved, the reader is always advised to seek a legal opinion from an expert qualified or knowledgeable about current law in the jurisdiction involved.

Ethics, at least in the tradition of Western philosophy, derives from two traditions: deontology and relativism. Each handles issues of competing ethical duties differently.

Ethical principles that flow from morals, religious interpretation, revelation, the instructions or commandments of the deity, Kant’s categorical imperative (a secular view of what must be done), or a collective cultural agreement on what is right and just (such as Confucianism) are called deontological. They tend to be absolute (such as “thou shalt not kill”) and therefore subject to fine parsing and reinterpretation when applied to complicated events in the real world (“... except in self-defense?”). This inevitably leads to a large body of theology or philosophical interpretation (such as the Talmud, the fatwas of Islamic scholars, or commentaries on the Bible in Christian tradition) and to competing schools of thought. Traditional medical ethics are deontological and are based on the following four principles:

- There should be respect for the autonomy of the individual (“autonomy”).
- Physicians should do no harm (“non-malfeasance”).
- Physicians should strive to do good (“beneficence”).
- All individuals have equal rights and responsibilities (“distributive justice”).

From these principles flow subsidiary ethical rules such as truthfulness, privacy, confidentiality, and fidelity.

Relativism, which is sometimes used in a pejorative sense, holds that ethical principles are meaningful only in relation to one another and to the facts of the situation. It accepts that there may be conflicting obligations and conflicts between legitimate but competing rights and responsibilities, forcing often uncomfortable decisions. Hard choices must often be made. Within relativism there are other paths, such as consequential ethics (taking into consideration the consequences and effects of applying the ethical principle) and utilitarianism (“the greatest good for the greatest number,” associated with the philosopher Jeremy Bentham).

In general, medicine has favored deontological ethics, but increasingly the deontological approach is proving inadequate. Physicians experience being caught between competing interests, constraints on how they can practice, limits on their ability to advocate for their patients, and ambiguous ethical dilemmas, regardless of their practice. Because occupational medicine places the physician in a direct relationship with parties other than the individual patient, potential conflicts of interest and competing ethical obligations often occur, and ethical issues may grow quite complex in OEM. Relativistic ethics is the only clear way to resolve these issues, but the nature and priority of competing obligations and rights must be clear. Occupational medicine has had many years to grapple with this problem. Medicine as a whole, at least in North America, is really just waking up to the reality that ethical duties often require balance.

In occupational medicine, much depends on the role of each of the agents or parties in the relationship. In occupational medicine, distinct agents can be identified by role:

- The service provider, who is the OEM physician or his or her designate. The service being provided may be a service of assessment and treatment, assuming responsibility for care, or of evaluation and recommendation without assuming responsibility for care.

- The client, who receives the service. This is usually the worker, who may or may not be a patient, in the sense of having a health problem, and there may or may not be a physician-patient relationship.
- The customer, who buys and in a sense “owns” the service. This is usually the employer (or a unit within a large employer) or the workers’ compensation carrier.
- Stakeholders, who may have a legitimate interest in the provision of care or the identification of the case. These may include regulatory bodies, appeals tribunals, suppliers who are concerned with use of their products (for example, chemical manufacturers and distributors involved in Responsible Care®), voluntary organizations, the injured worker’s personal physician, and others, including family members of an injured worker. Some of these stakeholders have a right to confidential information, outlined in law, and others do not.

In addition, the worker who is being provided care by the occupational physician may fill different roles, some of them at the same time. A worker who is well and who may be undergoing a periodic health surveillance program is not generally regarded as a patient—that term is usually reserved for someone receiving individualized medical services for a condition. A worker who has been injured or made ill is often called an “injured worker” (this term is regularly used in workers’ compensation circles) rather than a patient. A worker, whether injured or not, may be employed, self-employed, or unemployed (because “worker” describes as much the intent to work as employment). If employed, the worker is also an employee, which implies that the worker is bound by contractual obligations and policies that may influence his or her care (such as selection of the initial treating physician).

These are not distinctions without a difference. These distinctions matter. In the past, some occupational physicians believed that their true client was their customer, the employer, and that they did not

owe the worker disclosure or even consideration. They identified with management and saw no conflict in ethics, because they did not see themselves as having a physician-patient relationship with the worker; they considered the company to be their “patient.” No right-thinking occupational physician should think that way anymore. Clients have rights that should not conflict with customers’ rights if the system is working.

The physician should put the interests of the patient or injured worker first, but not at the expense of truthfulness or the potential to do significant harm. The corollary rights of the other parties must be respected and balanced, though the heaviest weight should favor the patient or injured worker. In case of conflicts, it is best to make the balance clear to each party and to document the situation in writing. This is so common that it can usually best be done on forms and confirmed by obtaining a signature to document that the worker was informed.

Occupational medicine practice and the bounds of ethics absolutely require clarity on when a physician-patient relationship exists and on making it clear to the worker when it does not. When the worker is being assessed and treated by the physician, a physician-patient relationship exists. When a medical evaluation is conducted solely for the purpose of a third party—such as a fitness-for-duty evaluation, disability evaluation, an independent medical evaluation, or periodic health surveillance—a patient-physician relationship does not normally exist, and if it does not, this has to be made clear to the worker. In situations where the client is not a patient, this should be made clear to the parties involved, and there is still an obligation for fairness, protection of confidentiality, and non-malfeasance (in the sense that the physician should not intentionally do harm but still has a duty to be truthful in reporting his or her findings).

In cases where the worker is also a patient of the physician in another context, such as serving as his or her personal physician, it is difficult and sometimes untenable to separate these roles. Understandably, there is also a trust issue and enough examples and stories of abuse for workers to generally prefer to seek the opinion of their

own physicians in matters such as disability evaluation or fitness to return to work. Some physicians assume that their role is to advocate for their patients' wishes, which may or may not be justified, rather than maintain arm's-length objectivity, as required by the occupational health or workers' compensation system. In so doing, they put themselves in an untenable situation, and their opinions may not carry much weight. That is why it is much better for occupational health services to be provided by qualified (and of course ethical) physicians who do not have an ongoing relationship with the worker in personal health matters. However, this is not possible in many situations, such as when there is limited access to qualified physicians. As in so much else in occupational medicine, this problem often needs to be managed on a case-by-case basis.

Some ethical issues are almost unique to occupational medicine, such as the appropriate management of periodic health surveillance and fitness-for-duty evaluations and disclosure of information. Others are common to the practice of medicine as a whole and are separate problems only as they must conform to the occupational health system, such as confidentiality of medical records. The majority of ethical problems that are of particular concern to the occupational physician are those that pertain to the behavior of the physician as confidant of the worker and as agent of the employer.

The physician who is seeing a patient on behalf of an employer or insurance carrier has an obligation to report to those parties such information as is directly pertinent to the employee's work capacity or work-related disorder and no more. The employee cannot enjoin the physician from conveying such information, nor can the employee enjoin the physician to fail to notify the appropriate government agency of a work-related injury or illness, as these are legal responsibilities of the physician. The employee has an absolute right to know the nature, name, and probable cause of a disorder that is found, and the physician has an absolute duty to inform the patient-employee of these findings and of what information will be conveyed to third parties. Under workers' compensation rules, however, the

employee may not necessarily be free to select any physician to treat a disorder beyond the first encounter with the expectation that unauthorized fees will be paid. When circumstances arise in which these rights and responsibilities create a conflict, something is wrong. The physician must then analyze the situation to determine who is out of line and stand firm in an unambiguous position, regardless of personal interest and sympathies.

Confidentiality of medical records is a major issue in occupational medicine, as it is in medicine in general. The general rule applies that the physical medical record is the property of who owns the file or the platform in which it is recorded, but that the information it contains belongs to the person who created it and is at all times available to the worker or patient. This leads to conflict when supervisors or human-resources managers attempt to access an employee's record. They are not entitled to do so. Disputed workers' compensation cases and litigation are subject to law, and the record must be supplied. At that point, control is often lost by the occupational health service. However, the OEM physician should always treat the record as confidential, just like any other medical record, securing it and limiting access to qualified health professionals. Although it is not clear that HIPAA (the American law governing handling of medical information) necessarily applies to occupational health or workers' compensation records, the OEM physician should not put it to the test and should have no reason to try; one should follow the same procedures as for medical records in general medicine.

Occupational health records are normally retained in the United States for at least thirty years (up to forty in some OSHA standards, such as for lead and arsenic) and for the duration of employment, as required by OSHA. (Employees who work for less than one year are not subject to this requirement as long as their records are given to them on separation.) As a matter of ethics, the records should be kept securely under the control of the OEM physician or the occupational health service during that time, or they may be transferred with the worker's consent to another party, including the worker's personal physician. Some state agencies and OSHA will also receive

and store records when an employer goes out of business, but normally the record becomes the property of the business that takes over ownership.

### **Codes of Ethics**

Codes of ethics for occupational medicine and for occupational health professionals in general emphasize seven basic themes:

- An obligation on the part of the employer and responsible occupational health professionals to provide a safe and healthy workplace environment for the worker
- An obligation on the part of occupational health professionals to maintain professional competence
- An obligation on the part of occupational health professionals to report problems and advise on solutions
- An obligation on the part of occupational health professionals and employers to maintain the confidentiality of the worker's medical records
- An obligation on the part of occupational health professionals to avoid conflicts of interest
- An obligation on the part of occupational health professionals and employers to avoid discrimination, intended or otherwise
- An obligation on the part of occupational health professionals and employers to maintain ethical standards

Obviously, it is much easier to apply and enforce these ethical standards on occupational health professionals than on employers.

The world standard in ethics for occupational health is the International Code of Ethics for Occupational Health Professionals of the International Commission on Occupational Health ([http://www.icohweb.org/core\\_docs/code\\_ethics\\_eng.pdf](http://www.icohweb.org/core_docs/code_ethics_eng.pdf)). The most recent revision at the time of this writing was approved in 2002 and is available in eight languages. The International Commission on

Occupational Health (ICOH) code has been widely adopted by other organizations (including the Association of Occupational and Environmental Clinics) and has been adopted as the legal standard of practice in Italy and Argentina. The American College of Occupational and Environmental Medicine was revising but had not completed its code of ethics at the time of this writing (2009). The Occupational and Environmental Medical Association of Canada has a short, streamlined code adopted in 1989—adapted from the Canadian Medical Association and the Ontario Medical Association—that is compatible with the ICOH code. The ICOH code will therefore be the basis for further discussion in this section.

The ICOH code covers all occupational health professionals, reflecting a shared set of values. This is also important in practical terms because of the confusion and the potential for exploitation that would result if different occupational health professions (physicians, nurses, hygienists, psychologists, ergonomists, and so forth) followed different rules on, for example, confidentiality. The ICOH code therefore applies to any professional who works in the domain of occupational health and overlays the standard of practice and ethics for the profession or discipline. In the case of medicine, the ICOH code guides the physician within the domain of OEM practice, but neither replaces nor subordinates medical ethics in general.

Principles of the ICOH code are laid out in a lengthy introduction, omitted here. Key provisions include the following:

- The highest goal of occupational health practice is protection and prevention.
- A safe and healthy workplace is the responsibility of the employer.
- The workplace should be changed; healthy working conditions should be provided; and work should be adapted to, rather than exclude, the worker.

- “Occupational health professionals should assist workers in obtaining and maintaining employment notwithstanding their health deficiencies or their handicap.” This reflects the emerging view that disability is a condition, not a characteristic, and that the way to manage it is to remove barriers to full participation of the worker to the extent feasible, and that a worker’s individual needs should be considered.
- Occupational health professionals are defined by their role, not by credentials or formal titles. This role comes with responsibility for workers’ health, which, once assumed, requires the practitioner to protect and promote workers’ health and commitment to a safe and healthy workplace.
- Discrimination is not acceptable.
- Occupational health professionals should be independent, free to give advice, adequately provided with resources, and allowed to practice according to the highest professional standards. (This provision implies but does not spell out a duty on the part of the occupational health professional not to practice at all if it cannot be done right.)
- Recognition that protection of workers’ health is primary, but that there are balances between, for example, protection of employment and protection of health (for example, when a worker has a condition that might present a threat to him- or herself or others but does not wish for it to be disclosed) and conflicts between individual and collective interests (in the same example, between an obligation to the worker and the obligation to prevent harm to other workers on the job).
- Recognition that much of occupational health practice is governed by law and regulation, and that there are also various International Labour Organization conventions and recommendations that are binding in countries that have ratified them. (Several of them require worker representation in occupational health services, which is not the norm in the United States.)

### ***Basic Principles***

Following the introduction, the text begins with a set of three basic principles:

- *The purpose of occupational health is to serve the health and social well-being of the workers individually and collectively. Occupational health practice must be performed according to the highest professional standards and ethical principles. Occupational health professionals must contribute to environmental and community health.*

This is a strong, cohesive statement, not a string of platitudes. The first principle requires high, not merely adequate, standards of practice, but recognizes a potential balance between the interests of the individual worker and the collective interest of fellow workers where they may conflict; this is different from the almost exclusively individual-patient-centered focus of conventional medical ethics. It lays the groundwork for the doctrine of balance among the interests of the parties in occupational health without considering such balance to be an ethical compromise.

- *The duties of occupational health professionals include protecting the life and health of the worker, respecting human dignity and promoting the highest ethical principles in occupational health policies and programmes. Integrity in professional conduct, impartiality and the protection of the confidentiality of health data and of the privacy of workers are part of these duties.*

Again, this may appear obvious, but it lays a foundation. The OEM physician (in this case) is being admonished to behave according to the highest standards of (in this case) medical practice and not to consider occupational health to be an area in which one can let down one's guard or compromise on behavior.

- *Occupational health professionals are experts who must enjoy full professional independence in the execution of their functions. They must acquire and maintain the competence necessary for their duties and*

*require conditions which allow them to carry out their tasks according to good practice and professional ethics.*

This principle puts the occupational health professional, in this case the OEM physician, on notice that it is one's own responsibility to determine whether his or her work is adequate, competent, and sufficiently well supported to be effective. The purpose of this principle is not to force any OEM physician to quit if he or she does not get all the support he or she wants. It is intended to require occupational health professionals to be careful about their relationships to ensure that their work is undertaken honestly and that they can follow their best judgment, rather than management dictates. Otherwise one ends up being merely a tool or mouthpiece, or a name to demonstrate compliance with local requirements, with no real authority. The principle also requires that occupational health professionals, in this case the physician, be diligent in preparing to practice in this field, so that, for example, a physician who markets his or her practice to employers in order to get occupational injury cases without learning anything about the workplace, work capacity, workers' compensation, or other important aspects of OEM would be considered an unethical practitioner.

### ***Duties and Obligations***

The body of the ICOH code then commences with a section on the duties and obligations of occupational health professionals.

1. *The primary aim of occupational health practice is to safeguard and promote the health of workers, to promote a safe and healthy working environment, to protect the working capacity of workers and their access to employment. In pursuing this aim, occupational health professionals must use validated methods of risk evaluation, propose effective preventive measures and follow up their implementation. The occupational health professionals must provide competent and*

*honest advice to the employers on fulfilling their responsibility in the field of occupational safety and health as well as to the workers on the protection and promotion of their health in relation to work. The occupational health professionals should maintain direct contact with safety and health committees, where they exist.*

This passage is straightforward. It should be noted that for employers of significant size, the ICOH code expects there to be joint (worker-management) health and safety committees or other mechanisms for the review, evaluation, and correction of workplace hazards and the assessment of safety performance, injuries, and incidents. American OEM physicians may have difficulty with this provision because such committees are uncommon in the United States (except where they have been written into a collective-bargaining agreement), but they are required by law in Canada. Employers may be reluctant to allow OEM physicians who work under contract to participate regularly in these internal company meetings. The OEM physician should at least ask to be on the distribution list for the minutes of the meetings, read them carefully, and follow up on any topic of medical concern.

2. *Occupational health professionals must continuously strive to be familiar with the work and working environment as well as to develop their competence and to remain well informed in scientific and technical knowledge, occupational hazards and the most efficient means to eliminate or to minimize the relevant risks. As the emphasis must be on primary prevention defined in terms of policies, design, choice of clean technologies, engineering control measures and adapting work organizations and workplaces to workers, occupational health professionals must regularly and routinely, whenever possible, visit the workplaces and consult the workers and the management on the work that is performed.*

This provision has become increasingly difficult for OEM physicians who work in the community and not as part of a corporate medical department or in a plant facility. It requires that an

effort be made to see the actual workplace and to talk to workers, both outside and inside the clinic, in order to better understand the realities of the workplace. This requires management cooperation, and the lack of such cooperation may be a sign that management will not provide the support and acceptance that the OEM physician needs to do his or her job. In some cases, it may be possible to visit similar workplaces elsewhere to gain a general understanding of hazards and jobs, but the heart of this provision is that the OEM physician should be familiar with the specific workplaces where the workers under his or her care are employed.

3. *The occupational health professional must advise management and the workers on factors at work which may affect workers' health. The risk assessment of occupational hazards must lead to the establishment of an occupational safety and health policy and of a programme of prevention adapted to the needs of undertakings [enterprises] and workplaces. The occupational health professionals must propose such a policy and programme on the basis of scientific and technical knowledge currently available as well as of their knowledge of the work organization and environment. Occupational health professionals must ensure that they possess the required skill or secure the necessary expertise in order to provide advice on programmes of prevention which should include, as appropriate, measures for monitoring and management of occupational safety and health hazards and, in case of failure, for minimizing consequences.*

The major thrust of this provision is that occupational health practice is not a solo affair to be assigned to the occupational health professional. It is management's responsibility and should be governed by a policy, as are other aspects of business and corporate responsibility. The employer needs to have a well-defined and explicit policy, and there needs to be a systematic program for carrying out that policy. If the employer does not have such a policy (a serious mistake, for many reasons, that can lead to big trouble for an organization), the OEM physician (or

other occupational health professional) should propose one, and it should be sound in principle and appropriate to the workers' needs. A policy might include a statement on the company's commitment to prevent injury and illness; a statement committing the company to meet or exceed compliance with all applicable regulations and laws; safe work practices (confined spaces, lock-out procedures, working alone or in isolation, and special needs of the industry); hazard evaluation and management (hazard communication and training); occupational hygiene surveys and monitoring; plant inspection; fire and security procedures; management accountability; first aid and access to medical care; early and safe return to work; training; incident investigation; reporting incidents and injuries; drug testing and substance abuse in safety-sensitive jobs; fitness for duty and work capacity (for example, compliance with the Americans with Disabilities Act); and training, at all levels, including drills. However, it is not enough to write the policy and hand it over to management, which may then put it on the shelf unread and unused. Management must be engaged in its development and must "own" it for it to be an effective document.

4. *Special consideration should be given to the rapid application of simple preventive measures which are technically sound and easily implemented. Further evaluation must check whether these measures are effective or if a more complete solution must be sought. When doubts exist about the severity of an occupational hazard, prudent precautionary action must be considered immediately and taken as appropriate. When there are uncertainties or differing opinions concerning nature of the hazards or the risks involved, occupational health professionals must be transparent in their assessment with respect to all concerned, avoid ambiguity in communicating their opinion and consult other professionals as necessary.*

It may not be obvious why this provision is included, but there is often a tendency in occupational health to wait for the

perfect solution instead of controlling the hazard quickly and expeditiously. This provision also calls on the OEM physician, in this case, to respect that there may be differing opinions regarding risk, but to be clear in his or her assessment and to consult other professionals, most often occupational hygienists, rather than guessing. This provision reinforces the emphasis throughout the document on prevention and on prompt action.

5. *In the case of refusal or of unwillingness to take adequate steps to remove an undue risk or to remedy a situation which presents evidence of danger to health or safety, the occupational health professionals must make, as rapidly as possible, their concern clear, in writing, to the appropriate senior management executive, stressing the need for taking into account scientific knowledge and for applying relevant health protection standards, including exposure limits, and recalling the obligation of the employer to apply laws and regulations and to protect the health of workers in their employment. The workers concerned and their representatives in the enterprise should be informed and the competent authority [government regulatory body] should be contacted, whenever necessary.*

This provision requires the OEM physician, in this case, to speak up and not let control of potentially serious hazards drop due to lack of management interest or action. However, it is also clear that the threshold for pushing the issue this far is an exceptional and substantial risk. Minor injury does not rise to this standard.

6. *Occupational health professionals must contribute to the information for workers on occupational hazards to which they may be exposed in an objective and understandable manner which does not conceal any fact and emphasizes the preventive measures. The occupational health professionals must cooperate with the employer, the workers and their representatives to ensure adequate information and training on health and safety to the management personnel and workers. Occupational health professionals must provide appropriate information to the*

*employers, workers and their representatives about the level of scientific certainty or uncertainty of known and suspected occupational hazards at the workplace.*

This provision requires transparency and effectiveness in risk communication on the part of the OEM physician.

7. *Occupational health professionals are obliged not to reveal industrial or commercial secrets of which they may become aware in the exercise of their activities. However, they must not withhold information which is necessary to protect the safety and health of workers or of the community. When needed, the occupational health professionals must consult the competent authority in charge of supervising the implementation of the relevant legislation.*

This provision acknowledges the legitimacy of trade secrets but holds the OEM physician, in this case, to a higher duty committed to health protection. Even so, it requires that local laws and regulations be consulted.

8. *The occupational health objectives, methods and procedures of health surveillance must be clearly defined with priority given to adaptation of workplaces to workers who must receive information in this respect. The relevance and validity of these methods and procedures must be assessed. The surveillance must be carried out with the informed consent of the workers. The potentially positive and negative consequences of participation in screening and health surveillance programmes should be discussed as part of the consent process. The health surveillance must be performed by an occupational health professional approved by the competent authority.*

This provision is not clearly written. Periodic health surveillance to screen for health effects has been more controversial in many countries (including Canada) than in the United States. Sometimes tests have been used that had little value or were even medically contraindicated (such as low-back x-rays). (See Chapter 5.) This provision requires that the methods of surveillance be evidence based. It also requires that participation be voluntary and based on informed consent such that a worker

can opt out, for example if a worker is concerned that his or her job might be at risk if he or she is found to have an adverse health effect, or if the worker just does not want to know. This provision would not apply to mandated surveillance, such as that required by OSHA standards, or drug surveillance for workers covered under the regulations of the U.S. Department of Transportation. If there is a legitimate need for health surveillance and it is not covered by a regulation, the procedure and the rationale for it should be documented thoroughly in a policy, and permission to carry out the surveillance should be included in the work contract. As always, seek the advice of legal counsel.

9. *The results of examinations carried out within the framework of health surveillance must be explained to the worker concerned. The determination of fitness for a given job, when required, must be based on a good knowledge of the job demands and of the work-site and on the assessment of the health of the worker. The workers must be informed of the opportunity to challenge the conclusions concerning their fitness in relation to work that they feel contrary to their interest. An appeals procedure must be established in this respect.*

In much of the world, fitness-for-duty evaluations are considered part of surveillance, hence the inclusion of both in this provision. This provision requires the OEM physician to disclose to the worker the results of surveillance testing and to explain their significance. The major part of it, however, outlines the approach to work capacity and fitness for duty (as presented in Chapter 18), and requires that the worker have some mechanism for appeal. In North America, medical conclusions on fitness for duty for new hires or return to work are not normally subject to appeal, although decisions based on those conclusions, being management decisions, can be challenged on legal grounds, as under the Americans with Disabilities Act. This provision requires an appeals mechanism, and fairness and transparency would dictate that the

conclusion be reviewed by a physician not involved in the first evaluation.

10. *The results of the examinations prescribed by national laws or regulations must only be conveyed to management in terms of fitness for the envisaged work or of limitations necessary from a medical point of view in the assignment of tasks or in the exposure to occupational hazards, with the emphasis put on proposals to adapt the tasks and working conditions to the abilities of the worker. General information on work fitness or in relation to health or the potential of probable health effects of work hazards, may be provided with the informed consent of the worker concerned, in so far as this is necessary to guarantee the protection of the worker's health.*

This provision is set in the context of pre-placement and periodic health evaluations required by law or regulation in many countries. However, its requirements also apply in North America, where practices differ. As required under the Americans with Disabilities Act (ADA) and Canadian law, the employer is not normally entitled to the diagnosis or any specific information on the medical condition of an employee or even details of the disability (although under ADA the employer may ask about the nature of the disability if the disability is obvious), except a conclusion regarding fitness for duty (expressed as "fit," "unfit," or "fit with accommodation/modification") and the accommodations or modifications required in the event of a disability. (This provision would also apply to certification of illness as grounds for absence by any physician; the "doctor's note" should not mention diagnosis.) Exceptions arise when disclosure of otherwise confidential medical information is required by law or regulation (as in workers' compensation claims), with informed consent, when disclosure would be in the worker's own interest (for example, if the worker became seriously ill at work), and when disclosure would be necessary for the public interest (for example, if the worker had a highly communicable disease). The first two are unequivocal, but the

latter two are relative and place the OEM physician in a difficult position of balancing interests. The standard for disclosure without the protection of law or consent would be very high: serious injury or risk to life.

- ii. *Where the health condition of the worker and the nature of the tasks performed are such as to be likely to endanger the safety of others, the worker must be clearly informed of the situation. In the case of a particularly hazardous situation, the management and, if so required by national regulations, the competent authority must also be informed of the measures necessary to safeguard other persons. In his advice, the occupational health professional must try to reconcile employment of the worker concerned with the safety or health of others that may be endangered.*

This provision acknowledges that the worker's right to employment must be balanced against the potential hazard to others if the worker's duties would endanger others. For the protection of others, "safety-sensitive" workers are held to a higher standard of fitness and reliability, and it is this standard that justifies more frequent periodic health surveillance and mandatory drug testing. However, the emphasis in this provision is on making the worker him- or herself aware of the risk to others, which gives the worker a chance voluntarily to request a change in duties. The provision sets a higher threshold for notifying management and government regulatory agencies; the situation must exceed the usual degree of risk in any workplace.

As in many ethical questions, there is also a legal dimension to this, which, insofar as it represents a social consensus on what is right, informs the ethical argument. In the United States, the decision in a highly influential 1976 California case, *Tarasoff v. Regents of the University of California* (551 P.2d 334), established a "duty to warn" for health care providers, which overrides confidentiality of medical records when there is a serious risk to others. Many states have laws requiring physicians to disclose information germane to public safety, such as a medical

condition that would interfere with safe driving, and others protect the physician who discloses against the will of the patient in such cases but do not require it. However, case law is thin. Canadian law and consensus on ethics accept the principle that confidentiality of medical information is not absolute and that public safety may override it, but this consensus sets a high standard, such as a risk of serious injury or death that is imminent and concrete, not distant or theoretical. Thus it would appear that at least in North America society has decided that confidentiality should be protected to the extent possible, but not at all costs, and that serious injury and death are unacceptable costs.

The ethical principle is that neither the confidentiality of medical information nor the duty to warn of possible risk is absolute. They must be weighed, and the only sure balance favoring disclosure is when the consequences of failure to disclose information are unacceptably high to society. To the maximum extent possible, the OEM physician should avoid such situations, such as by persuading the worker to disclose the information voluntarily (and documenting the effort); by making a judgment of “unfit” for duty; or by seeking guidance through regulation, contract language, and strong employer policies.

12. *Biological tests and other investigations must be chosen for their validity and relevance for protection of the health of the worker concerned, with due regard to their sensitivity, their specificity and their predictive value. Occupational health professionals must not use screening tests or investigations which are not reliable or which do not have a sufficient predictive value in relation to the requirements of the work assignment. Where a choice is possible and appropriate, preference must always be given to non-invasive methods and to examinations, which do not involve any danger to the health of the worker concerned. An invasive investigation or an examination which involves a risk to the health of the worker concerned may only be advised after an evaluation of the benefits to the worker and the risks involved. Such an investigation is subject to the worker's informed consent and must be performed according*

*to the highest professional standards. It cannot be justified for insurance purposes or in relation to insurance claims.*

Biological tests, in this context, mean clinical, laboratory, bio-monitoring, or personal sampling tests (such as biological exposure indices). To be acceptable, they must conform to the characteristics of a well-performing test as outlined in Chapter 5, particularly with respect to predictive value. The provision discourages or prohibits tests that have not been scientifically validated (such as visual contrast sensitivity as a test for toxicity), that do not have well-defined outcomes (such as neuropsychological testing), that are not demonstrably related to work exposures or capacity (such as pelvic examinations for female employees), and that have poor predictive value (such as color vision tests as an initial screen for solvent toxicity). The provision also prohibits tests that carry an unacceptable risk (such as low-back x-rays) or inconvenience the worker (such as 24-hour urine collections). Every so often, someone invents a new test and tries to use it for screening workers without waiting for appropriate validation and acceptance by the scientific community. (A real, recent example was a proposal to test urine for asbestos fibers, which made no medical sense.) As a practical matter, this means that only time-tested, reliable, and well-validated tests should even be considered for screening, periodic health surveillance, and fitness-for-duty evaluations. Otherwise, the risk of discrimination because of false positives or negatives is too high.

Genetic testing is particularly sensitive. In addition to providing inappropriate information on individual susceptibility and health risk to employers, genetic testing can be used unfairly to exclude workers from a workplace with a hazard, even though the focus should be on controlling the hazard.

13. *When engaging in health education, health promotion, health screening and public health programmes, occupational health professionals must seek the participation of both employers and workers in their design and*

*in their implementation. They must also protect the confidentiality of personal health data of the workers, and prevent their misuse.*

This short provision is particularly important because it insists on the involvement of workers in the means for their protection. It treats occupational health and health promotion as a matter of equity, in which workers have the right to a say in how their health and interests will be protected, without denying that primary responsibility always rests with the employer.

The ICOH code has provided occupational health and safety professionals around the world with consistent guidance for appropriate behavior and has established a global standard of practice. It cannot solve every conceivable dilemma, but if its principles are followed diligently and documented, the OEM physician will be on firm ground.

## **THE WAY FORWARD FOR OCCUPATIONAL HEALTH AND SAFETY**

Occupational health and safety (OHS) is like a cork in a sea of economic activity, bobbing up and down as the waves pass under it, never completely sinking but carrying little weight, and always at the mercy of forces much larger than its pull on gravity. As the U.S., Canadian, and world economies pass through a wrenching transition, it is time to consider where occupational health and safety may be headed and how the forces at play just now, at the time of writing (2009), are shaping it. The future of OHS will define, to a large extent, the future of OEM.

Occupational health and safety has been seen in many different ways in its long history. The field has passed through eras in which it has been perceived to be primarily an issue of public health, of production efficiency, of labor-management relations, of corporate responsibility, of loss containment, of liability control, and now of human-resources management. What is needed now is a new conceptualization

that strips occupational health and safety of its accreted social and political agendas and returns it to a practical field in public health and medicine. This does not mean backing off from a commitment to workers' health protection as a human right. It does mean that this right should be served by a professional system that is stable, objective, and separate from political agendas. (The system should not be neutral, however, because it exists to protect workers.)

Society as a whole has never placed a high priority on safe workplaces as a public health goal or value, preferring to deal with issues of productivity and loss control and to avoid the human rights dimension of occupational health protection. Occupational health and safety is generally perceived as something to be managed, like warranties on equipment or maintenance on a plant. Occupational health and safety functions are perceived as an expense of production, to be kept under control but not likely to yield a return on investment (although they can!). Identification and control of highly specific occupational hazards may not hold much appeal in terms of benefit and cost in the new economy, especially because so many employers are persuaded that the major industrial hazards, such as asbestos, belong to the past. Occupational health services are easily lost in the shuffle in health care policy and management. That is why reform of occupational health and workers' compensation should not be folded into the mainstream of health care reform; there, it will always be an afterthought.

### **New Challenges**

A major challenge to OHS professionals is to create an attitudinal change in the workplace. OHS is considered in industry and by most workers to be an add-on responsibility, not a normal part of doing business. For OHS professionals to achieve their potential, occupational health and safety must be an integral part of the mainstream.

It is instructive to compare interest in the environment with interest in occupational health and safety. The two are clearly

related in some people's minds but are very distinct to others. Environmental quality is clearly an issue of public concern; occupational health and safety is perceived by the public as a much narrower issue, of concern only to selected groups of workers and special interests. Environmental quality is always highly visible; occupational health and safety is recognized only in times of crisis or in dramatic cases. Environmental quality is an end in itself and a desirable enhancement of the quality of life; occupational health and safety is a means to an end, because it facilitates productivity and removes obstacles to efficient operation. Small wonder that the quality of the environment is a powerful shaping issue in our society, but the quality of the working environment goes through cycles of neglect and rediscovery.

Since about 2000, the United States has entered a new and rather regressive phase in occupational health and safety. The emphasis today in industry, government, and organized labor is "back to basics"—a renewed emphasis on the essentials, dispensing with more-sophisticated approaches. There is a feeling that occupational health and safety may have become too complicated and technical, and that the basic commitment may need to be reviewed. The result has been a turning inward to reexamine basic issues, rather than reaching outward for new ideas and initiatives. It is an odd phase of rediscovery because it fails to build on what has gone before.

As a result of the election of 2008, there have been fundamental changes in political leadership in the United States that have raised interest in occupational health and safety and environmental protection. However, the onset of the global recession has left government, major corporations, labor, and public opinion with a sense of confusion. The forward momentum expected in occupational health and safety may be over before it started. On the other hand, environmental protection remains high on the national agenda, and emerging green technology makes it easier to align environmental priorities with fiscal incentives and economic reconstruction.

In the economy that emerges from the current recession, occupational health and safety will probably become a production and sustainability issue for some industries, mostly those on the economic margins, and a health services management issue for most. Employers are likely to give higher priority to issues that affect more workers, that affect the workers in higher-value-added jobs, and that involve greater monetary losses or potential productivity gains. These are the issues that relate to low-risk workplaces, such as ergonomic hazards and psychogenic stress—issues that increasingly dominate the occupational health and safety agenda. Occupational hazards of a physical and chemical nature are likely to get worse, or they will at least continue to present management problems in manufacturing sectors that are already under intense economic pressure, as well as in small enterprises that are least able to manage them well.

Occupational health and safety professionals in the new economy may have to learn, as a business proposition, to deal with occupational health primarily through issues embedded in productivity, personal health insurance, disability management, adjudication processes, and wellness. In order to keep employers' attention, the argument may need to be made on the basis that a dollar invested in occupational health and prevention of injury pays back much more in recovered costs than most investments.

In the new economy, some occupational hazards may be novel; some may be redistributed; and some may require specialized technology to control, which has been the historical experience in, say, the semiconductor industry. However, most of the major hazards are likely to be the same ones seen today in sectors such as construction. Serious risks will increasingly be concentrated in certain industries and groups of workers, among them disadvantaged and special populations. The practical management problem for occupational health and safety in the future, as in the past, will be to deal effectively with hazards that are fragmented and often irrelevant to the economy, and to deal with problems that are too often buried in aggregate statistics and nearly invisible from the outside—problems for which there is

therefore little political will to do anything. But after an era of incentives for underreporting, apparently declining occupational injury rates, and competing priorities for management attention, will there be the will to do anything?

## **The Way Forward**

What is the way forward? If society is to attain and, more importantly, sustain effective efforts in occupational health and safety, supporting a new priority for OEM, there need to be certain changes:

1. Consistency. OHS programs need long-term stability to thrive.
2. A stronger OHS infrastructure. Occupational health and safety needs a broader base in industry, government, labor, and academia.
3. Independent voices. New ideas usually come from small groups working intensively outside of day-to-day responsibilities. There needs to be respect for the role of university programs, union-sponsored institutes and centers, and worker-led organizations (such as the Committees for Occupational Safety and Health, social scientists, technical consultants, trade associations, and professional organizations), because these are society's think tanks for new ideas. The first rule of brainstorming for new concepts is to be wide open to new ideas and to evaluate them later. That is how practical, effective, and innovative ideas emerge.
4. Mainstream commitment. Special award programs and such are very good and have an attention-getting effect for a few years after their initiation. What really counts, however, is the integration of occupational health and safety into production and management: criteria for promotion, performance reviews, operating procedures, bidding on contracts, contractor performance, collective bargaining, fiscal projections, and operating

- permits. The rest becomes window dressing after the first few years.
5. Disregard the distinction between safety and health. The professionals in each professional sector should be talking to one another all the time, regardless of their background and training.
  6. Change the paradigm. Occupational health and safety is by rights a public health function, not a privilege or a negotiating position.
  7. Return to the intellectual basics, but not to kindergarten. The blueprint for effective OHS regulations has already been written. Society does not need to reinvent it. The fundamental features of the Occupational Health and Safety Act of 1970, for example, are not the cause of OSHA's disappointing performance. Reform may be needed in OSHA, but the basic legislative philosophy has not failed.
  8. Open the system. Occupational health and safety thrives when experiences are shared among professionals and especially among employers, and when management listens to its own workers. Occupational health and safety always falters when employers and professionals maintain strict isolation, refuse to share information, and take a top-down approach to the management of safety, acting by directive to their workers instead of learning from their workers' experiences.
  9. Openness to new ideas. Canada and the United States represent two quite different traditions in occupational and environmental health protection. The United States itself has fifty experiments in state regulation and state-level workers' compensation. Much can be learned by comparing them and by examining other national models, such as the United Kingdom's Health and Safety Executive, which to most observers has been conspicuously more successful than OSHA.
  10. Continuous improvement. For occupational health and safety standards as well as for environmental standards, a national policy

of continuous improvement makes sense. Standards are now proposed and adopted as if they are permanent and as if re-evaluation is an exceptional event. Standards do not stand for all time: they are superseded as more information becomes available, technology advances, and the economic cycle permits. Just now, the introduction of green technology and infrastructure reconstruction constitute one of the few optimistic lights in an otherwise dismal economic agenda. This is an historic opportunity to make things better. Why not institutionalize continuous progress and stop fighting it?

11. An end to environmental standards battles. Battles over new standards are, for the most part, ultimately delaying tactics. How much less disruptive economically and politically might it be to accept small and frequent incremental changes as natural steps, advancing a little more at every technological advance and opportunity without fighting individual battles over the documentation and justification for large-increment jumps in standards?
12. A unitary view of occupational and environmental health and medicine. It has been a consistent theme of this book that occupational medicine and environmental medicine are sister disciplines that in this age are merging, though that process is not yet complete. The next step in bringing these fields together may be to start thinking of occupational medicine as a subset of environmental medicine, a medical field concerned primarily with the built environment of the workplace and of sustainability. Sustainability in this sense is not environmental or ecological sustainability—preserving resources for future generations—nor is it economic sustainability—supporting the same harvesting (or exploitation) year after year. There is also a form of cultural sustainability, based on a shared vision of stewardship that keeps culture and values intact. Part of this vision of sustainability is that health should not be compromised for gain and that a sustainable society protects the health of those who make its wealth and continuity possible.

## RESOURCES

- Guidotti TL. Environmental and occupational health: a “critical science.” *Int Arch Occup Environ Health.* 2005;80(2):59–60.
- Guidotti TL, Hancock T, Bell W. Physicians and environmental change: what is our role and why should anyone listen to us? *Perspect Biol Med.* 1998;41:591–604.
- Ravetz J: Towards critical science. *New Sci.* 1971;51:681–683.

## NOTEWORTHY READINGS

- Canadian Medical Association. *Health, the Environment and Sustainable Development: The Role of the Medical Profession.* Ottawa: CMA; 1991.
- Dickens C. *Hard Times.* Mineola NY, Dover, republished 2001. [Classic novel, first published in 1854 in London, and generally considered to be Dickens’ most severe condemnation of the Industrial Revolution. It satirizes the mindset of the “Strict Father” as described by Lakoff and in the text.]
- Guidotti TL. Preventive medicine, public health, and the environmental movement. *Am J Prev Med.* 1991;7 :124–125.
- Heilbroner RL. *The Great Ascent: The Struggle for Economic Development in Our Time.* New York: Harper and Row; 1963. [Robert Heilbroner had already predicted, in the 1960s, that economic growth and production would be limited by social restrictions on external costs, such as environmental degradation and occupational health and safety issues.]
- Lakoff G. *Moral Politics: How Liberals and Conservatives Think.* Chicago, University of Chicago Press, 2002, 2nd ed. [First edition, published in 1996, was subtitled “What Conservatives Know that Liberals Don’t”.]
- McCally M, Cassel CK. Medical responsibility and global environmental change. *Ann Intern Med.* 1990;113:467–473.
- Ravetz JR. *The Merger of Knowledge with Power: Essays in Critical Science.* London: Continuum International; 1989.
- Ronchi D. The quality of working life movement: part I, history. *Employee Relations.* 1981;3(1):2–6.
- Task Force on the Implications for Human Health of Global Ecological Change. *Human and Ecosystem Health: Canadian Perspectives, Canadian Action.* Ottawa: Canadian Public Health Association; 1992.
- Weber M. *The Protestant Ethic and the Spirit of Capitalism.* Mineola NY, Dover, republished 2003. [Classic treatise in sociology discussing the influence of values and faith on economic behavior. First published in German in 1905.]
- Wells DM. *Empty Promises: Quality of Working Life Programs and the Labor Movement.* New York: Monthly Review Press; 1987.

# **26 GLOBAL OCCUPATIONAL AND ENVIRONMENTAL HEALTH**

Occupational and environmental medicine (OEM) physicians, like the employers and workers they serve, are engaged on one level or another with the global economy by virtue of being in business. The rise of modern globalization, and of interdependence through trade and finance, connects most enterprises of any scale, directly or indirectly, and affects most workers in developed and developing countries. Occupational and environmental health problems also play an important role, generally unappreciated, in economic development, both as a drag on the national economy as it grows, a drain when it falters, and as a factor that always tragically limits the participation in modern life of those whom it affects when there is disability.

The word “global” is increasingly used to describe a broader range of issues and relationships, within and among countries, including those that affect minorities and vulnerable groups (such as indigenous peoples). In an increasingly integrated world, the word “international” is insufficient to describe common threats, risks, and responses, and it is also an insufficient word to describe the levels of cooperation and communication between peoples. “International” also implies, if only grammatically, bipartite or multipartite relationships between and among nations or countries, rather than the whole of health around

the world, whereas “global health” as used today also applies to issues within countries and to vulnerable populations such as internally migrating populations, whether refugees or economic immigrants. Health issues are rarely unique to a particular country or society, but neither are they identical among different countries. With the exception of access to health care, health issues are rarely demarcated by a national border. The study of commonalities and differences in health and its determinants from one national community to the next, and how to change disparities in health, defines global health.

A small number of OEM physicians work mainly or exclusively for multinational companies or for international organizations in the global arena. Many have responsibilities for operations outside North America and for workers, including executives, who travel. Most provide services to the local workers of domestic employers, which are profoundly affected by globalization and international trade. Fewer physicians—but a significant number of them—work on environmental health issues around the world, usually through research. Whatever their level of involvement, an understanding and appreciation of global occupational and environmental health helps the OEM physician to see the big picture.

Global occupational and environmental health is the world perspective on determinants and problems of health relevant to OEM and other occupational and environmental health professions.

Global occupational and environmental health has many dimensions, among them:

- The role of occupational health in economic development worldwide
- The role of environmental health in economic development worldwide
- Management of occupational health in foreign subsidiaries and operations
- Health issues at international borders, associated with movement of people

- Travel medicine, as it applies to workers and management
- International agencies and organizations

## **ECONOMIC DEVELOPMENT**

OEM physicians are concentrated in developed countries, of course, and are in short supply in developing countries. The OEM physician in North America is likely to be concerned primarily with occupational, rather than environmental, health in developing countries because demand for services is driven by the need to serve and address issues relating to foreign operations of the employer. It helps to understand the development process in a general way in order to visualize how occupational health services fit into development and play a role in economic development. The same overview helps in understanding the role of environmental health, which is less frequently a responsibility of the OEM physician with responsibilities in developing countries.

Attitudes toward occupational and environmental health are embedded in the social frameworks and public health systems of their societies and countries, as discussed in Chapter 25. Every country's approach to occupational and environmental health reflects its history, economic development, and culture—particularly attitudes toward risk—and is affected to some degree by its diversity. However, some generalizations can be made. One is that very few societies make occupational health a priority in their affairs (one exception is Finland). Its role in economic development is almost completely overlooked. Attitudes toward the environment are much better defined, however. As a society grows richer, it tends to become more risk-averse and insists on more protection. As a consequence, development tends to bring along with it, at least in later stages, an intolerance of personal risk and an insistence on public protection, whether from crime, pollution, or economic insecurity. Environmental protection also becomes a priority, because of a desire to prevent environmental health problems and a growing appreciation for environmental quality, both for its own sake and as an economic advantage. Within a given

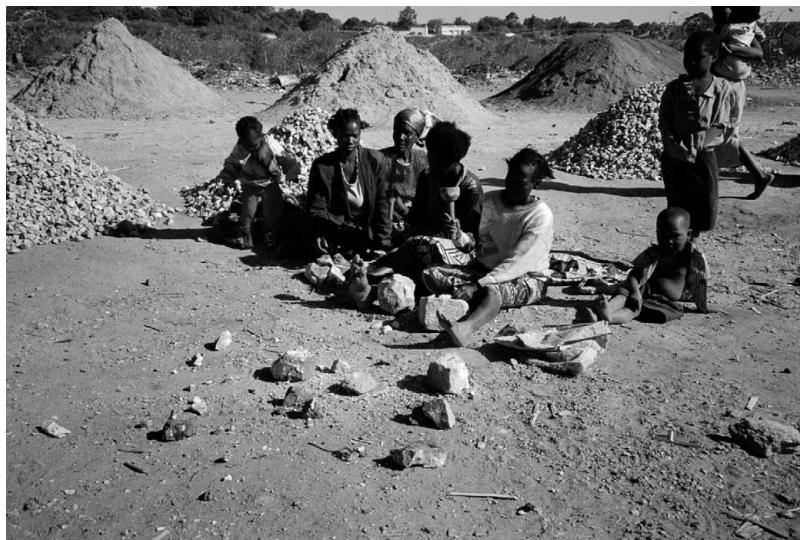
level of development, however, there seems to be almost no correlation between what priority a society places on environmental quality and the priority given to occupational health and safety, perhaps because occupational health is out of sight.

## **The Basics**

Before undertaking any discussion of economic development, it is critical to think of the human faces behind it. Economic development too often reflects statistics and generalizations. The reality is that economic development is all about real people struggling to survive under difficult circumstances and how, how often, and how long the lives of some are made easier and the lives of others are made worse; because in any economy, there are winners and losers, and development is largely a matter of creating conditions to make more winners than losers.

Figure 26.1 shows a glimpse of the reality. The family in the photograph is working on the outskirts of Lusaka, the capital of Zambia. They are using a piece of scrap steel to break rocks into gravel. This is how they earn their living, every day, in the hot sun. They keep their children with them because there is no other way to look after them. They sell the gravel they make every day to a man who then sells it in bulk to a construction supplies dealer, who had a ready market when this photograph was taken because Lusaka at the time was experiencing a mild surge in building. If something goes wrong, and a rock chip flies into someone's eye, there is no health insurance or money for medical care. If the able-bodied adults who are strong enough to break the rocks are injured or become ill (in this country with a high rate of HIV/AIDS), who will earn the income for the family or look after the children? There is no safety net. This family is at an extreme end of poverty, trapped in an economic vise, and barely hanging on. Their survival as a family, if family members become sick or injured, is an open question. And yet this family is relatively fortunate because at least it has something to sell, a customer to buy the product, and income to share. Zambia once had a thriving economy

based on copper, with a per capita income approaching European countries. After a series of reversals, including nationalization of the industry and collapse of commodity prices, the national economy crashed, thrusting Zambia into dire poverty. Commodity prices only recently recovered, due to demand from China. Were it not for this, there would be no housing construction and no demand for gravel. Both are ultimately dependent on economic trends on the other side of the world. Economic development is about people. Economic theory is an essential guide to how to create wealth that can be shared, and economic data are essential to test the reality. However, all too often the human face of economic development is lost in a fog of theory and statistics. At its core, economic development is about creating wealth for a purpose: so that all people can live their lives in a manner that frees them to pursue their own goals and aspirations, play the roles that they can play in their families and in society, and reach their own potential as human beings.



**Figure 26.1.** Family by the side of a road on the outskirts of Lusaka, Zambia. They are breaking rocks into gravel to sell for building material.

All countries are developing, in the sense that their economies are constantly changing into something else. The economy and level of technological and social development (a process often called “modernization” or “urbanization” in development shorthand) do not stand still. The United States, for example, moved in one century from an agriculture-dominated country of rural residents and small towns, through a robust manufacturing-dominated economy, to a service industry-dominated urban society, still with a large manufacturing sector, now caught, due to the world economic crisis, midway in transition to something resembling a greener economy dominated by information and innovation. Canada has experienced something similar, but together with Australia, it remains an anomaly among highly developed countries because extractive industries remain critical to the economy. In other words, economic development is not simply a process of countries ascending a stepwise ladder, getting richer in the process, until they are industrialized. It is a constant process of evolving that can go backward (as in the case of Zambia) as well as forward, and it does not stop with industrialization. Development also comes with an environmental cost that is rarely counted and evaluated.

Globalization is an economic regime that involves worldwide trade and a high degree of specialization by countries or regions in economic activities in which they have a natural advantage, such as natural resources (as in Australia), land (as in Brazil), or abundant intellectual talent (as in India). Countries that are highly dependent for their advantage on natural resources such as copper (Congo, Zambia) or crops such as coffee (Ethiopia, Guatemala, Ivory Coast) face great economic uncertainty when commodity prices fluctuate, which they always do over time. Another problem with resource-based relative advantage is that it creates a strong incentive for over-exploitation of resources, such as forest products in Indonesia and Thailand and overfishing in the world’s oceans. The relative economic advantage of a country is not permanent. In the Arabian Gulf, Bahrain and Dubai (which is only one of the United Arab Emirates, but the one forced to develop more commercially) are running out

of oil and so have diversified their economies. A relative advantage occurs when countries pass through a phase of high manufacturing capacity and low wages, making them exceptionally competitive for manufactured goods for export and for outsourcing production, as is the case in China in recent years and Vietnam today.

Outsourcing production creates jobs and raises wages in the countries doing the manufacturing, but the manufacturing step is very sensitive to costs and is not where the big returns lie in the value chain. Therefore, once wages rise to a certain level, the same countries try to outsource to other countries, as is seen historically with the movement of outsourced low-margin business in Asia from Japan to Taiwan or from Hong Kong to coastal China and now to interior (and poorer) China, Vietnam, and Cambodia. Unfortunately, as business moves from one country to the next, occupational hazards tend to follow the outsourcing in the form of unsafe work practices, unenforced occupational health standards, obsolete equipment, and untrained and unaware managers and workers.

Until 2009 it was relatively easy to categorize the winners and losers in the world economic system. The score was kept in terms of “gross domestic product” (GDP, the value of goods and services produced by the country), which is universally acknowledged to be deficient as a single metric for productive wealth and is used for that purpose anyway. GDP has many drawbacks, not least that it does not take into count, in some form of debit, the environmental and health costs of production, and it treats nonproductive and counterproductive costs—such as health care for preventable illness, tobacco consumption, and occupational disorders—as if they were productive economic activities. GDP is sometimes qualified by measures such as income distribution (the ratio of income of the highest earners to that of the lowest earners), which reflects social characteristics that relate to social cohesiveness and stability. The most important of these is the Human Development Index, which incorporates scales for per capita GDP, longevity, and standard of living (income taking into account “purchasing power parity,” which is an adjustment for the cost of living).

The onset of the world recession has completely upset this neat way of thinking. Because of the initiating factors of the recession, economic instability began and was greatest at first in the most developed and richest economies. Countries that have been relatively isolated from global markets have tended to do better, at least in the early phases of the crisis. It is probably not an exaggeration to say that textbooks of economic development are now generally obsolete and will need to be rewritten. Although it is not clear that the current economic crisis represents the decline and fall of capitalism, as many radical critics believe, it is clear that the triumph of capitalism was declared prematurely. Two decades after the collapse of Marxist-Leninist socialism (the Soviet-style Communistic variety), capitalism itself is now in crisis. To be fair, the form of capitalism that is now in trouble was a largely unregulated, market-driven variety with many abuses, and it may be argued that it was a corrupt form that only needed reform. Whatever the verdict of history, what will come next in many countries is likely to be more regulated and *dirigiste* (state directed, as in the French system) than the old laissez-faire variety. If the economic future does involve more government oversight and regulation, environmental and occupational health may get more attention.

Terminology is important in global development work. Specific jargon is used by each international organization or country, such as the United Nations, the U.S. State Department, the U.S. Agency for International Development, and each of the numerous other governmental and nongovernmental organizations involved in global health. Some concepts and terms are common to each, however, and their strict meanings are not always familiar to Americans, who tend to use the words “nation,” “country,” and “state” more or less synonymously. A “nation” is properly a group of people with shared history, culture, and language, sometimes called a “people.” However, not all such groupings are nations. Nations largely define themselves, such that even closely related peoples with a shared history or the same or almost identical languages, such as Czechs and Slovaks or, tragically, Hutu and Tutsi in Rwanda, may consider themselves to be

separate nations. Within nations there may be tribes, a term that implies a self-recognized unit with a traditional, cohesive structure based on kinship ties and usually organized by clans. A “country” is a political entity that has defined borders. In most countries, the apparatus of governance has two parts: the “state” and the “government.” The state is the corporate entity that represents and rules the people; the government is the specific apparatus of power that does the governing and can be changed without changing the state. There is a sharp distinction between governmental organizations (usually called ministries or departments), nongovernmental organizations (universally called “NGOs”; environmental NGOs are commonly called “ENGOs,” although some of them think the term is pejorative), and UN or other “international” organizations or agencies. Very few NGOs are concerned with occupational health.

The terminology that describes developmental levels is constantly shifting. Countries that have developed an industrial or modern and urbanized infrastructure (in the sense of urban amenities and culture, not necessarily big cities) can be called industrialized, developed, advanced (the term is now out of favor), or more developed (the latter abbreviated MDs), terms that generally correspond to the level of income (as measured by gross domestic product, or GDP, per capita) of member countries of the Organization of Economic Cooperation and Development (OECD, twenty-seven member states), which includes western Europe, Japan, the United States, and Canada. Those that have achieved this level of infrastructure and income recently, such as Malaysia and Thailand, are called “newly developed countries.” Which countries merit this label changes with each generation (newly developed countries of the previous generation included the “Asian tiger economies” of Hong Kong, Singapore, South Korea, and Taiwan).

Developing countries are those with infrastructure inadequate to need, a smaller level of economic activity, lower income, and often, but not always, fewer strong social institutions that are national in scope. The term “less developed countries” (abbreviated LDCs) is now out of fashion. There are no hard-and-fast criteria for the category,

and the category is very heterogeneous: the World Trade Organization (WTO) allows countries to declare themselves “developing” as they see fit. Some developing countries have done well in terms of income and social services, such as Costa Rica and Botswana, and others have not, such as Bolivia. A few developing countries, such as Cuba, have experienced the special history of state socialism and Soviet-style Communism, whereas others, such as Georgia, Armenia, and Azerbaijan, have emerged from that history; China took a different path altogether. “Least developed countries” are those low-income countries that are recognized by the WTO as the poorest and most struggling, including Bangladesh, much of sub-Saharan Africa, Haiti, Nepal, and many isolated island countries. The recent world economic recession, however, has turned economic predictions for many of these countries upside down, because economic growth in some “less” developed countries, such as Togo, has been proportionately much greater than in stronger economies.

Every developing country has its own story. Economic profiles say little about how most countries arrived at their situation. Emerging economies are those undergoing a transformation, either breaking through to a middle-income level through a conventional development path or coming back after setbacks, such as those recovering from a history of Communism and the aftermath of the collapse of a Soviet-style system. The BRIC countries (Brazil, Russia, India, and China) are four large economies that play an increasingly important role in the world economy, but as a group they have nothing else in common. It is quite possible for countries to experience catastrophic reductions in income and economic development, as has been experienced in recent years, at different levels, by Argentina, Zambia, and, most recently and tragically, Zimbabwe. Some developing countries, such as Nigeria, Venezuela, and Kazakhstan, have oil resources; in development theory there is much talk of the “oil curse,” because the commodity tends to distort national economies by preventing diversification and introducing opportunities for corruption, such that many of these countries fail to achieve their potential. Some countries inherited an infrastructure from colonization that proved useful

in the postcolonial period (such as Ghana and India), and others were devastated by colonialism (such as Congo). These examples demonstrate that although there are similarities in the economics of developing countries, there are also many exceptions and special cases. China, the largest country of all, is also the greatest exception of all, having experienced in its long history almost every setback conceivable, including colonialism, yet having developed rapidly by following unique policies.

“Middle-income countries” are developing countries that are well along the path to development and have achieved much more than subsistence levels of income for their people, but that are below OECD income levels, such as Mexico and Iran. The World Bank and, for most purposes, the World Health Organization prefer to label countries by income level: low income (such as Pakistan, Myanmar, Laos, Mali), lower middle income (Indonesia, Colombia, Philippines, Moldova, Honduras, Ukraine), upper middle income (Belarus, Turkey, South Africa, Libya, Chile), and high income (New Zealand, Slovenia, Macao, Bahrain). However, this can be misleading. Most of the world’s poor at present (2009) live in lower-middle-income countries. Some higher-middle- and upper-income countries are very small and depend on somewhat artificial economic foundations, such as tourism and financial services. The essential magazine for following world economic trends is the *Economist*, a British news-magazine universally read in development circles, which simply calls countries “rich” or “poor” and otherwise avoids making too many generalizations.

Economic development proceeds along a political as well as an economic path. This must be understood by anyone working in global occupational and environmental health. National priorities may be very different in some countries: health is not necessarily the highest priority. Ministries of finance are always the most powerful government departments in their countries, and issues of occupational and environmental health, although not environmental protection, tend to be invisible to them. It is not unusual for the ministry of health to be low in status compared to other government departments or for the

minister of health to be given the job for political reasons and to have no particular interest in health matters. Ministers of health rarely see environmental health and especially occupational health as falling within their mandate, although they may accept them as part of a comprehensive public-health program. Environmental health is often seen as a matter for the ministry of environment (particularly in countries with strong ENGO movements) and occupational health as a matter for the ministry of labor. Both of these ministries, in turn, tend to consider health to be outside their core mandates and almost always lack the required infrastructure in public health to do much anyway, with the possible exception of clean water and vector control.

It is important to distinguish between “population health risk” and “personal health risk,” because the former is what is measured at the country level and the latter is what is experienced, especially by workers, and what drives occupational health issues at the local level. The difference is something like the difference between public health and medicine. Population health status reflects the health of the population on average or for the community as a whole, as reflected by the frequency of public health problems and the level of health according to various indicators related to personal risk and behavior, and how these health indicators are distributed. Personal health outcomes are individual illnesses or injuries that result from the interplay between the causes of the injury or illness and the risk factors present in the individual. For example, heart disease is more likely in people who smoke and who have a family history of smoking. Population health factors such as cholesterol in the diet are important for the population as a whole, but may or not be important for the individual. When such a person has a heart attack, it is a personal illness with specific implications for that one person; it may not have any particular implications for the risk of other people, especially if that person ate a different diet or had different life habits. Therefore, population health is not just the aggregate of personal health characteristics in the population, but a description of how the population behaves as an intact, organized community with different levels and subgroups. Measures of population health status represent

the risk of poor health for the majority of people and the current state of poor health for a minority of sick people, who, at any one time, need access to health care. The total illness burden on society as translated to the level of population health is affected by how the society defines illness and chooses to accept or to act to prevent a burden of illness that results in disability. (This is one reason that the standard measure of the burden of disease in international comparisons has become the calculation of “disability-adjusted life years” [DALYs].)

### **The Development Cycle**

Contemporary models of development have re-evaluated the fundamental relationships between economic expansion and social progress. The influential work of Amartya Sen has raised awareness on the true value of economic development, which is that it allows the individual to engage in social roles and personal aspirations and is therefore a community and a personal good. This line of reasoning places great emphasis on human rights and social development as an integral part of economic development. However, the predominant view among economists is still that economic growth is an end in itself.

Economists often speak of the “economic development cycle.” This is the idea that economies progress stepwise through a predictable sequence of events that begins with a fundamental reliance on agriculture and a population living primarily in rural communities. This is called an “agrarian” economy. As productivity increases, more surplus crops can be sold and traded, but the basis of the society remains agrarian until it either reaches a point where it can support a process of industrialization, or industrialization is introduced. In the nineteenth century, industrialization (to a limited degree, so as to preserve most of the market for exports from the colonial power) was introduced by colonial powers. In the twentieth century, it was often financed by foreign aid and bank loans. In the twenty-first century, more industrial growth is the result of private debt, foreign direct investment, and domestic investment.

Industrialization yields products and services in anticipation of an excess of revenue over cost, which can then be invested back into the economy as capital for further development. Manufacturing also lowers the cost of finished goods and makes them more widely available at lower costs, eventually coming within reach of the average person, who becomes a consumer. However, economic development does not stop there. It is now known that the economic cycle progresses into an economic model based on services and information. At each step along the way, there are dislocations, disruptions, and instabilities, as well as advances in creating wealth. “Globalization” refers to trade and specialization of manufacturing and economic activity based on comparative advantage, such as access to resources, special talent, or low wages. The globalization of the manufacturing economy has also resulted in the export of numerous hazards in industry to countries poorly equipped to control them, among them asbestos. It has also resulted in the unsustainable exploitation of resources to meet a global demand, as in deforestation in Southeast Asia and overfishing in the world’s oceans.

The history of how rich countries developed is an imperfect guide to what developing countries are experiencing. Developing countries develop differently than they did a few decades ago; for example, they skip steps in infrastructure development that are no longer relevant. Railroads are no longer considered critical to economic development; highways have taken their place. Technology—the cell phone even more than the Internet—has made a huge difference in the economics of developing countries. Because of the cell phone, citizens of developing countries are able to communicate easily, can learn about market prices for their commodities before they bring them in for sale, and can work through much larger social networks. Economists and decision makers know much more today and have access to much more information. It is not necessary to repeat mistakes of the past.

On a national level, finance ministers and development economists tend to believe that industrialization requires investment in production first and that once wealth is created, it can be invested in social

goods such as improved health, worker protection, and protecting the environment. The fallacy is that by this theory, social goods, including health and work security, are amenities—nice but not essential. From the worker's point of view, however, a disabling injury usually means slipping back into poverty and social marginalization. Another way of looking at the problem is that basic occupational health services bring the workers and their families into full participation in the development process by protecting their lives, health, and income. The reasoning is the same as that underlying the theory of work capacity (see Chapter 18). In fact, the enhancement of health on a population level means the creation of opportunity and capacity, and provides the individual worker and his or her family with security, the potential for growth in income, and the capacity to do more in the community that merely subsist.

The key insight in new theories of development is that health and a decent income give a person the capacity and autonomy to function in society. This links the idea of economic development with human rights, which can be seen as a cluster of values that define the life and freedom of action of a citizen of the society. Of particular importance is the freedom and autonomy enjoyed by women. Social change becomes much more rapid and crosses generations more readily when women are empowered to engage in civic roles, to educate their children with different values, to obtain an education, to participate as partners in family decisions, to work in jobs out of the house or that are not necessarily traditional for women, to make reproductive choices, and to manage money. Gender equality shows a strong correlation with economic development.

The balance between population and available resources is another critical dimension in economic development. In the 1970s, there was much concern about overpopulation and the risk that increasing numbers of people in the world would outstrip the resources available. The imbalance between population and resources may seriously aggravate social and economic problems at a particular moment in time, but the trend over time has not been as dire as predicted. As it happened, the relationship turned out to be much more complicated,

with population growth slowing and resource availability being profoundly affected by technology. That present trends, even though not as extreme as predicted, cannot continue indefinitely is clear, but the history of population issues over past decades shows that the issues are not a simple question of an arithmetic progression involving the availability of food and an exponentially increasing number of mouths, as famously described by Thomas Malthus. For one thing, increasing affluence and longevity tends to reduce population growth markedly by reducing fertility (in a demographic, not a biological, sense), providing at least a partial self-correction. In terms of fairness, access to the resources needed for an acceptable life is an issue of human rights.

The contribution of technology as a root cause of pollution and resource depletion was the other major theme of the environmental movement in the 1970s and, like population growth, has had to be revisited because the effects are much more complicated than projected at that time. Technology allows much greater efficiency and opportunity, which offset or mitigate its negative effects from the introduction of, for example, new products or processes that present risks. (See Chapter 25 for a discussion of “critical science” and the critique of technological progress.)

Urbanization (sometimes called “modernization,” although this term has fallen out of favor) is the process of bringing together infrastructure, a critical density of people, political organization, a wage-based economy, and vastly greater social networks. It also incorporates the application of and access to technology in the lives of individuals and in society, during the initial years of agriculture and manufacturing. Urbanization as a social process is, obviously, concentrated in cities, but lines of communication, transportation, and migration also reach back to and affect rural life. There may be rapid incorporation of urban habits, lifestyles, tastes, and ideas into the life of villages and rural communities, often starting with fashion and music. Rural poverty is worse when the village is remote from towns, commodity prices are unstable (as they usually are), and infrastructure is poor (which it usually is). Rural areas also tend, as a rule, to be conservative

and traditional. Urbanization brings in new ideas and new ways of thinking.

### **Globalization and Trade**

The relationship between globalization and trade has been identified by many public health activists and some economists as a broad issue of concern, both as a public health issue and as part of a larger critique of globalization. The principal issues are:

- Trade agreements and the legislation to support them may subordinate national legislation on health and environmental standards.
- Trade agreements often do not acknowledge environmental protection, environmental health risks, workers' rights, or occupational health; or they manage them through weaker side agreements, as in the case of NAFTA.
- Financial policies imposed on countries in financial distress (historically by the International Monetary Fund, IMF) may reduce the ability of the population to purchase food and health care or to provide care for young children. (The IMF has reviewed and changed its policies.)
- National health systems or services that involve subsidies of government financing may be ruled an unfair trade advantage. (The issue was raised in the negotiations of NAFTA with regard to Canada's health care system.)
- Products imported from countries with lax inspection standards may present a risk to the public. (Imported toys that are contaminated or painted with lead have been a particular problem in North America.)
- Pressures to reduce costs and maintain high margins may result in exploitation of workers, unsafe working conditions, and adulteration of products (such as the addition of melamine to milk in China in order to fraudulently increase the apparent protein content).

- Patent protection and enforcement of intellectual property rights may limit availability of drugs and medical care in poor countries.
- The failure of the Doha round of trade talks, which involved reducing barriers to export of agricultural commodities from poor countries to rich countries, has locked many countries out of free trade and into poverty. (Doha, a city in Qatar, is where the agreement was negotiated.)
- Globalization has put intense pressure on many poor countries that deplete their resources in order to supply demand, especially if there is global competition. Deforestation is a common example.

The evidence for some of these charges, such as the perverse effects of the Doha round and deforestation, is strong. The role of globalization in some other issues is less clear.

### **Occupational Health**

Occupational health protection is considered to be a “consumptive cost” by employers, a cost that is incurred by the employer for little or no gain, instead of as an investment that yields future benefit. The benefit is often difficult to quantify because it is expressed in terms of disability and health problems that have been avoided, not as a tangible return on investment. Employers often take the view that because their role in the economy is to create wealth, not to distribute benefits, it is not their job to protect the health of workers. This point of view ignores the ethical and social benefit of worker protection, but it finds some justification in the prevailing view in development economics that rising wages and economic development, on their own, do more than health interventions to improve health conditions.

On the other hand, the cost of health care for injured workers is low in developing countries, and workers are usually easily replaced. There is little economic incentive to invest in health protection,

safety, or health services for injured workers. Although the costs of providing health care and measures for occupational health protection are minimal on an international scale of value, they are still seen to be an extra cost of production, particularly if an employer is in a competitive international market.

The primary health care system, where it functions, must accept the burden for occupational injuries and illness for most of the developing economy. There is usually no organized insurance or workers' compensation system, although employers may voluntarily pay some compensation to families of injured workers. Most enterprises in all economies, especially developing economies, are small and cannot support their own health care systems.

Most developing countries have sparsely developed occupational health and safety services. Such services are most likely to be available in the private sector because they are usually not given a high priority in the health care system. The World Health Organization (WHO) and the International Labour Organization (ILO) have therefore called for a commitment to "basic occupational health services" (BOHS), to be provided within the primary health care system through the training of health care providers (mostly physicians and nurses) and the addition of occupational health and safety personnel to primary care clinics and hospitals: a primary care level of occupational health, emphasizing primary prevention and hazard control, but integrated into the primary health care system. There is no question that BOHS is critically important to providing essential occupational health care to all workers, but whether it can be folded into a primary health care system, with its own priorities and limited resources, is still an open question. Prototype BOHS programs have succeeded, but most examples evaluated to date have been in China, which is not representative of other developing countries.

Lost income, in the form of wages that cannot be earned because the worker is disabled, can be devastating to the worker and his or her family, but in developing countries income levels are low to begin with. Because one worker easily replaces another, this disability-related loss is not perceived as an economic liability to the total

economy. Over time, however, disability-related loss begins to become a drag on the economy. The problem becomes more acute and governments pay more attention as income levels rise and the losses become more apparent. Most national social security systems (such as Mexico's "Seguridad social") started at this stage, with coverage for the employed population in middle-income occupations and their dependents, at a time when the country had achieved at least lower-middle-income status.

As the economy develops, the cost of health care quickly rises. The burden of occupational injuries and illnesses increases health care costs and reduces productivity. When disability pension schemes, where they exist, are added, the cost becomes conspicuously higher because injured workers are younger than the disabled in the general population, whose disabilities are more often associated with aging, and therefore each injured worker is more costly to treat and support over his or her remaining lifetime if he or she cannot work.

### **Child Labor**

Child labor is a huge humanitarian concern with great costs to society and the children themselves. However, child labor is difficult to eradicate because many families in some developing countries depend on the income. In most agrarian societies, children traditionally work in the fields and are an economic asset on farms. Seen through agrarian and village values and traditions, working outside the home does not seem abnormal, particularly where schooling is not valued highly. Children become economic liabilities when the economic level increases and attitudes change; then they are expected to stay in school. These changes in society do not happen uniformly or smoothly.

Historically, child labor was a mainstay in the new factories and mines of the Industrial Revolution and required a major social movement and landmark legislation to first control it (mostly by reducing work hours) and then effectively bring it to an end, which happened in the nineteenth century in Britain and in the 1920s in Canada as well

as in several progressive states (especially Illinois and Massachusetts) in the United States. Child labor was controlled and then mostly ended in the United States as a whole in the mid-twentieth century, largely due to the work of a federal agency called the Children's Bureau. It took longer to end child labor in the United States because several early laws against it were declared unconstitutional on the grounds that the federal government was not deemed at the time to have authority over working conditions, only interstate commerce. Even today, child labor, without appropriate limits and controls, continues to be a problem, particularly in agriculture. Fatality statistics for the United States continue to record work-related deaths in workers seventeen years of age and younger, indicating that some young workers are still working in risky occupations.

In developing countries especially, children are employed in the informal economy, where wages are not recorded, no taxes are paid, and no standards are enforced. This tends to conceal the problem from view and cover up abuse. Families often send the child away to live with an employer as a sort of apprentice or indentured servant. A special problem occurs with socially disrupted or migrant populations and separated children or orphans. In countries with a high prevalence of HIV/AIDS, children who are orphaned are often left to support themselves. For girls, especially, this usually means domestic work, but sometimes prostitution is the only apparent option for earning enough money.

The costs are huge to society and even greater to the children themselves. Children who work instead of going to school for a full day lose education and miss opportunities. They are also more susceptible than adults to occupational health risks. The ILO estimated in 1997 that as many as a quarter of a billion children were working in jobs that presented some hazard or risk. Young bodies are more susceptible to hazards, and children injured at a young age may be disabled for life. Children, even if physically capable of the work expected of them, are impulsive and do not understand the nature of hazards or the need to work safely; they do not have the knowledge to protect themselves or the authority to insist that an employer protect

them. The family, which may be far away, cannot effectively protect the child.

Child labor also leads to abusive situations because of the power relationship between adults and children, particularly if the children are employed outside the family, and because of the control over the child and his or her family exerted by money. Sometimes the family will turn a child over to an unrelated adult for money—virtually selling the child—expecting to receive an income from the child's work or, less commonly, in order to reduce the number of mouths to feed, with the expectation that the child will be taken care of. Girls are especially vulnerable to abuse because of gender discrimination in many traditional societies and because of the greater potential for sexual abuse and forced prostitution. At an extreme, the abuse can lead to child prostitution, crime, and virtual slavery. Children have even been kidnapped and forced into military service as child soldiers, with accompanying drug use, atrocities, and violence intended to keep them belligerent toward outsiders but pliant to authority.

The United Nations Children's Fund (UNICEF) monitors and documents trends in child labor and analyzes root causes in order to support efforts at prevention. It also administers the UN Convention on the Rights of the Child. ILO and UNICEF are particularly concerned that efforts to end child labor do not cause even more harm to children. UNICEF has compiled case studies where this has happened. Most involve dismissal and abandonment of the children when employers begin to feel pressure. The ILO sponsors the International Programme for the Elimination of Child Labor. One of the mainstays of the program is Convention 138, on the Minimum Age of Admission to Employment and Work (1973). Another is Convention 182, on the Worst Forms of Child Labour (1999), through which ILO seeks to eliminate the worst abuses first and then progressively raise awareness and change attitudes on child labor as a whole over time. In 2000 the United Nations adopted the "Palermo Protocol" (Protocol to Prevent, Suppress, and Punish Trafficking in Persons, Especially Women and Children); other conventions and protocols

that deal with sexual abuse and illegal migration in general also apply to children.

## **Environmental Health**

Environmental degradation is a predictable consequence of the development process in the absence of controls or a mature and inclusive civic culture. Environmental degradation is not exclusively the result of development, of course, and serious environmental health risks, particularly those involving water, are always present before economic development begins and can appear at any level of development. Clean water is critical for health, and control of contamination by human wastes is the most critical problem of all. Air pollution, in general, is more closely tied to the urbanization process and reflects the level of development of infrastructure, including the number of cars and trucks on the road and how energy is produced.

The most obvious impact of environmental degradation is illustrated by the severe air and water pollution encountered in many of the world's new megacities. Environmental degradation can also be seen as a cluster of issues that reflect potential adverse effects on health (which are usually difficult to prove); reduction of biodiversity and habitat (trading short-term economic gain for long-term destruction of potentially renewable resources); depletion of nonrenewable resources; limitations on future land use; risks to agricultural productivity and food supply; and diminished appeal for tourism, trade, and quality of life. Environmental degradation restricts future economic options, which is generally what gets the attention of ministries of finance.

At one time, the predominant attitude was that protecting the environment was something that only rich countries could afford to do, so developing economies tried to get rich first before making an effort to protect the environment. The global increase in awareness of the environment for its own sake—because of media attention, the risk of ecological catastrophe, and the desire to exploit commercial opportunities—has changed that, and now most governments at least

nominally support environmental protection, although they do not all follow through with support for that protection. Most societies are less interested in environmental health than in environmental quality. This is because in the political and social agenda of most countries today, environmental quality issues are increasingly seen as critical to the development of a sustainable economic base (and often tourism), but environmental health is seen as old-fashioned, particularly with respect to clean water, and tends to be seen as a by-product of an earlier generation of progress, particularly with respect to air pollution. These attitudes are not particularly rational, but they are deep, and they influence political priorities and decision makers. During the severe recession that is ongoing as this is being written (2009), more attention has been given to the prospects of economic recovery through rebuilding infrastructure and public investment to support a sustainable, “green” economy. This is new, at least on the present scale.

Environmental health, however, does not usually benefit from the attention paid to environmental quality as a whole. Relationships between human health and ecosystem conservation, biodiversity, and wildlife protection exist, but they are indirect. The constituencies and incentives to protect the rain forest and to protect human health from water in urban slums contaminated with feces are usually very different, and the former is generally much more appealing to politicians and business interests. (It should be noted that in recent years there has been a paradigm shift in the private sector, and more business interests are shifting to at least acknowledge the benefits of environmental protection, often seeing it as the basis for new industries such as ecotourism. It is also hard to sustain a tourist industry when there are periodic outbreaks of diarrhea.)

Issues of environmental health in developing countries are dominated, reasonably, by the need for a secure and clean water supply. Beyond that, the agenda is diverse and driven in large part by the concerns of the community as they understand the environment, not necessarily any objective standard. ENGOs play a disproportionate role in setting the environmental agendas of many emerging or

developing countries. For example, during the reconstruction of Eastern Europe following the collapse of the Soviet Union, much attention was paid to environmental issues that might have seemed marginal as health risks by North American standards, but that had been heavily publicized by ENGOs. One such issue involved opposition to a high-temperature incinerator for toxic wastes in a situation where no other disposal option was available, creating a bigger risk of dumping and uncontrolled release.

## **FOREIGN OPERATIONS**

OEM physicians who serve as medical directors in their organizations may have responsibilities for managing occupational and environmental health in foreign operations. For those companies based in the United States, Canada, the European Union, or another developed country, the corporate infrastructure for occupational and environmental health is normally already in place and functioning, and the challenge is to apply the same standards as in the home country while being responsive to the local situation. However, as business changes, it is increasingly common for large corporations to be truly multinational and for companies based in emerging and developing countries (such as Brazil and India) to own American subsidiaries and even parent companies. Then the challenge becomes to harmonize upward to the highest standard whenever possible over time.

Few multinational corporations have a simple corporate medical department anymore. More often, these departments include some responsibility for environmental health as well as other responsibilities: often product liability, sometimes safety (which is more often centralized), and occasionally security. The manager, often but not always an OEM physician, must, with a limited staff, simultaneously oversee and monitor the provision of medical services at the local level, manage corporate liability and compliance with policy, troubleshoot problems, anticipate and prepare for emergencies, and work closely with management to ensure that occupational and environmental concerns are reflected in corporate decisions, some of which

(such as new plant design) are not primarily about health. In addition to all this, there must be time to cultivate working relationships with other senior managers and to learn the culture and policies of the organization.

Managing foreign operations presents a number of challenges:

- Policy development
- Compliance with standards and best practices within the parent company
- Compliance with local standards and best practices, especially by audits
- Organizing an occupational health service that conforms to the corporate organization and is still responsive to local needs
- Identifying service providers at the local level
- Managing health issues for visitors
- Managing endemic disease in high-risk locations

Although mission and vision statements are important in every large organization, they are particularly important in defining occupational and environmental health policy in far-flung multinational corporations. Essential in all such policies is a commitment to comply with all applicable regulatory requirements, but there must also be a clear commitment on the part of the employer to occupational health and safety and environmental health protection as values. Without a clear statement of intent and unambiguous policies that support enforcement, local operations invariably go their own way, and compliance becomes inconsistent. This places the organization at extreme risk, for several reasons. The employer without tight control over local operations may inadvertently violate local regulatory laws, may have difficulty with quality assurance and exports, may lose control over occupational health protection and workforce health, and may face serious reputational damage, especially if the company is seen as applying different standards away from home. Strong and

well-monitored policies that honor a commitment to occupational and environmental protection are therefore fundamental to managing foreign operations.

A convenient approach to ensuring consistency of standards is to adopt well-recognized global standards, such as those of the International Organization for Standardization (ISO—the acronym is a pun on the prefix *iso-*, meaning “same” in Greek); the British Standards Institute’s OHSAS 18000 (which has applied for ISO recognition, but this has not yet been granted); the voluntary ILO Guidelines on Occupational Health Management Systems (ILO opposes ISO adoption of OHSAS 18000); or DEKRA Certification (particularly popular in German-speaking countries), depending on the function. The largest multinational employers often have internal occupational health and safety standards that are applied throughout the organization, rather than relying on local regulations. These are normally chosen to be lower than the lowest regulatory standards. In the past, they often conformed to standards recommended by the American Conference of Governmental Industrial Hygienists (ACGIH), but increasingly the harmonized standards of the European Union are being adopted.

The single most important tool for ensuring consistency in compliance with standards, corporate policy, and best practices is the audit. Audits should be undertaken at reasonable intervals, such as every few years, and should, ideally, examine outcomes (reportable injuries and illnesses, fatalities, near misses); reporting relationships; performance indicators; facilities and resources; budget; personnel; qualifications of professional staff; quality of case management; measures for quality assurance; training; adherence to procedures; disaster preparedness; availability of technical assistance for mitigating hazards; and actual performance in key functions: occupational health and safety hazards in place, environmental emissions and effluent, safety hazards, fire and hazardous materials management, engineering controls, management of personal protection, and housekeeping. In practice, the audit has to conform to the scope of the management unit or department, but it can uncover many gaps, and this information can be used to propose needed changes to management.

The occupational health and safety management structure must conform to the general pattern of organization of the company, or it will be out of step with general management. A company may have regional or national medical directors or a centralized corporate medical department that tries to do it all by managing hired local physicians. Which structure is preferable depends on how the corporation is organized and where it is operating. This may change over time, sometimes abruptly. For example, for most of its history Shell Oil was among the most decentralized of large companies, but in recent years it suddenly set out to restructure itself to a more traditional organization. Management of occupational and environmental health must follow.

Finding trained occupational health and safety professionals and environmental managers can be difficult in developing countries. One approach is to bring in expatriates to do the work, but this is often unsustainable over the long term. Expatriates should be partnered whenever possible to local talent so that the expertise can be transferred. Such an arrangement ultimately expands the talent pool in the country, raises awareness, and may even create new career opportunities in health and, particularly, safety. There is a strong temptation to “raid” academic programs for faculty talent, but this may have the unfortunate effect of destroying the program. A better option is to provide material support for whatever academic programs exist and to hire the students they produce.

Inevitably, a business visitor—such as a customer, government official, or corporate official from a partnering company—to the home country or to a hosted meeting held in a third country will fall ill or seek medical care. In general, there is no problem with the company providing emergency care, but it is always advisable to carry insurance to cover such situations. It may be very difficult for the company to stop providing care to a valued contact after the initial emergency is over, but the terms of the insurance policy set limits, making this easier. Some visitors will request medical care for themselves or their families as a quid pro quo of doing business. The OEM physician must be very careful about complying with such requests.

The Foreign Corrupt Practices Act of 1977 carries tough criminal sanctions against any employee or agent of an American company who provides or offers “anything of value” to a foreign official to influence decisions or to direct business. Similar legislation has been adopted by other OECD countries, including Canada. Bribery of foreign corporate officials is also against the law in the United States and Canada. Elective medical services are very high in value and clearly fall under these laws.

Foreign operations are sometimes located in zones of endemic disease that threaten employees, the local community, and business travelers to foreign operations. The most common serious situation of this nature occurs with operations in locations where malaria is endemic or where HIV/AIDS infection is prevalent. Where the local infrastructure is lacking, the employer may have to be proactive in establishing health services for the local population. Malaria illustrates this problem well and is a severe problem in several tropical regions where, for example, oil and gas are produced. It can best be controlled by the application of a series of partial measures for primary prevention (vector control, insecticide-treated mosquito netting, and long-sleeved shirts and long pants, as well as DEET and chemoprophylaxis for visitors), none of which are complete defenses but which together greatly reduce the risk. Some cases will occur, so access to effective diagnosis and treatment is required (diagnostic testing, first- and second-line chemotherapy) to prevent disability (particularly from falciparum malaria). If the local health care system cannot support malaria management programs, the employer may have to. Oil companies are in a position to do this, but many other industries are not. Likewise, HIV/AIDS and related problems such as infection with tuberculosis represent a major challenge in southern Africa—a challenge that has been met by a concerted and sustained campaign of prevention, medical treatment, and aggressive case management by the South African mining company Anglo American. Anglo American has emerged, of necessity, as a global leader in workforce health and disease management.

Ultimately, however, the only sure way of monitoring foreign operations is to visit. The OEM physician who assumes responsibility for widely distributed foreign operations should normally plan on spending 30 to 50 percent of his or her time traveling in order to do so.

## **BORDER HEALTH AND MIGRATING POPULATIONS**

Issues of border health and immigration are as critical within North America as anywhere in the world, and they are of particular interest in global environmental and occupational health. This is unusual in global health as a whole, which has traditionally not concerned itself with health issues in the United States and Canada (with the possible exception of aboriginal health issues) and has generally considered Mexico to be too developed, on a relative scale, to be of real interest.

Global health as a field has an informal specialization in migrant populations that is mostly concerned with refugees—who by definition are outside their country of nationality—and internally displaced persons. This often involves camps and involuntary population movement across international borders, and the driving issues are usually political or related to warfare or civil unrest. In general, border issues have not been of much interest otherwise in mainstream global health, with the exception of the U.S.–Mexico border. That particular border has attracted serious scholarship in global health, and in many other fields, because of its unique characteristics and because for so many years it was the most stark example to be found of countries at different levels of development in intimate contact. In part for this reason, this border area has experienced a plethora of health-related studies, demonstration projects, and commissions, mostly devoted to infectious disease, health promotion, and access to health care.

Global environmental and occupational health has a very different perspective, which places more emphasis on economic refugees and the problems of coordinating environmental and occupational health where two very different systems come together.

## Border Health

Occupational and environmental health along international borders presents many special issues. Because of the similarity in legal tradition, level of development, and economic organization, managing health issues along the U.S.-Canada border is mostly a matter of respecting the differences in authority and legislation (as covered in Chapter 24). The United States and Mexico are rather more different and at very different levels of economic development. Therefore, as a practical matter, the principal border health issues in North America requiring active management are those along the U.S.-Mexico border.

It should be noted that throughout this book, “North America” is used to refer to the United States and Canada, without reference to Mexico. Obviously, Mexico is part of North America geographically, but culturally it is part of Latin America, and the Spanish word *norteamericano* refers to American (some Mexicans prefer *estadounidense*, in recognition of the fact that they also share the continent and consider themselves *americano*) and, usually, to Canadians also.

The U.S.-Mexico border extends for 2,000 miles between San Ysidro, California, and Tijuana, Baja California, in the west and Brownsville, Texas, and Matamoros, Tamaulipas, in the east. It is the second-most-transited border in the world, after the U.S.-Canada border. There are four American states to the north of the border (California, Arizona, New Mexico, and Texas) and six Mexican states to the south (Baja California, Sonora, Chihuahua, Coahuila, Nuevo León, and Tamaulipas). The border area (often called *la frontera*) by convention (agreed to in the 1983 “La Paz” Agreement between the two countries) extends 100 kilometers, or roughly 63 miles, on either side of the international border itself. It is an area of intense cultural mixing, where English and Spanish are used interchangeably and daily life blends American and Mexican customs. Much of the area is sparsely populated, but there are several large, interfacing cities on opposite sides of the border, particularly San Diego and Tijuana, the two cities of Nogales (Arizona and Sonora), El Paso (Texas) and Ciudad Juárez (Chihuahua), Laredo (Texas) and Nuevo Laredo (Nuevo León), and

the larger cities of the Rio Grande River Valley (Texas and Tamaulipas). There are also numerous indigenous communities on both sides, most of them very small.

The border is also an area of unusual contradictions, where rich cities (including some urban areas on the Mexican side, which has a higher per capita income than the rest of Mexico) abut impoverished communities, including poor towns and Indian reservations in the United States and *colonias* (informal shantytowns) in Mexico. It is also an area beset by interrelated social problems, including high crime rates, drug use and smuggling, illegal immigration, trafficking in people, and high unemployment. Most of these problems affect both sides of the border and have proven rather intractable over the years. At the same time, economic development has been led, especially on the Mexican side, by the North American Free Trade Agreement (NAFTA, 1994) and by considerable cross-border local trade. Americans seek Mexican specialties, cheaper prices on high-value products (including pharmaceuticals), and low-cost services. Mexicans patronize stores in the United States for lower prices and a greater variety of merchandise.

As in most arid regions, the principal environmental health problem along the border is water supply and quality. Much of the U.S.-Mexico border is defined by the Rio Grande ("Rio Bravo" on the Mexican side), which has a very low flow and is inadequate to supply the growing population. There is a constant risk of fecal contamination due to sewer line breaks and spillover during flooding in many border communities. Substandard housing is a serious problem on both sides of the border. Air pollution levels have tended to abate over the years, especially with the closing of a smeltery near El Paso that had been the cause of local lead contamination for many years. Heavy pesticide use in agriculture, especially in the Imperial Valley (California), Mexicali Valley (Baja California), and Rio Grande Valley (Texas), has made the area a "hot spot" for pesticide toxicity and has raised issues of chronic toxicity. Toxic waste from local industry has created a serious local disposal problem, with several incidents of dumping, illegal discharge into water, and illegal transport into Mexico.

Some Mexican consumer products and folk medicines have been found to contain toxic levels of metals (especially *greta*, which contains lead). In the early 1990s, a cluster of cases of neural tube defects in children born in Matamoros raised concerns over pollution levels in water and possible exposure to toxic substances, such as solvents, during pregnancy. The issue was never resolved.

In 1965 the government of Mexico introduced a program to stabilize employment, promote economic development, and reduce economic migration to the United States in the border area. The plan facilitated the establishment of assembly plants and small factories, called *maquiladoras*, in duty-free zones such that American companies could bring unfinished products and components over the border, have them assembled by Mexican workers who were paid lower wages, and then have the assembled products returned to the American side of the border, again duty free on the U.S. side, for incorporation into the finished product and distribution. From the beginning, it was highly successful and attracted many companies. Over the four decades following, the *maquiladora* system came to employ over 1.2 million workers (one-third of Mexico's manufacturing workforce) and made northern Mexico more prosperous. It also created a trained and reliable workforce that provided a platform for economic development. However, the revenue generated by the *maquiladoras* was highly cyclical and depended almost completely on the U.S. economy. Employment tended to swing wildly between boom and bust, most spectacularly in 2001, when *maquiladora* employment dropped by almost a quarter within the year. Even so, the economy remains stronger and wages are higher in the border area than in most of Mexico.

The current outlook for *maquiladoras* is uncertain because wage levels on the border have risen out of competitive range for low-margin assembly and low-cost production work, with such work now being outsourced to China. On the other hand, productivity (per worker) has continued to rise and the workforce, after forty years, is now experienced, well trained, and adaptable, while still demographically young. The *maquiladora* sector is still competing successfully in more high-value products, such as surgical instruments,

and where proximity to the U.S. market is important—either because transportation costs are high relative to the value of the product, as is the case with durable goods such as appliances, or because product styles change rapidly, such as denim pants. Production requiring protection of intellectual property is also undertaken in the *maquiladoras*, because Mexican legal protection is stronger than in other countries with more competitive wages.

Despite the economic success of the *maquiladora* model, it has been very controversial. Many critics accuse it of exploiting Mexican workers, which is not surprising given that wage disparities are so obvious when so close in geographic proximity. Although occupational health protection is strong in Mexican legislation, enforcement is often lacking, creating a vacuum for compliance that companies could exploit in the *maquiladora*. Documentation of elevated rates of occupational injury and illness is hard to find for the region, and there are many anecdotal and unconfirmed reports. It seems clear that occupational health protection has not been as stringent in Mexico, but how this may have affected morbidity and disability rates is not documented.

Many government-sponsored binational and multilateral organizations are active on the border, most in a coordinating rather than an executive or operational role because of the fragmentation of authority. Because of the priority placed on the region, the Pan American Health Organization maintains an office in El Paso, Texas, to deal with U.S.-Mexico border issues. The U.S.-Mexico Border Health Commission, established by agreement in 2000, is a bilateral council of government officials with the objective of addressing shared health issues. The Border Legislative Conference is a forum for state legislators in both countries with the goal of sharing solutions and coordinating legislation. There is also a U.S.-Mexico Border Governors Health Table, established in 1980, and a U.S.-Mexico Border Counties Coalition, established in 1998. These official bodies have a similar agenda: emergency preparedness (terrorism being the priority on the U.S. side, natural disasters on the Mexican side, especially where there is a flood risk), control of communicable diseases, and strengthening infrastructure; environmental health has not been as much of a focus, and there has been essentially

no activity in occupational health. Despite this impressive structure of coordinating bodies, effective coordination has been elusive.

The U.S.-Mexico Border Health Association, established in 1943, is the professional organization devoted to issues of border health. It is headquartered in El Paso.

## **Immigration**

Immigration into the United States comes from all nations, but Mexico is the single largest source of immigrants, as it has been historically. The distribution of new immigrants to the United States, including those from Mexico, is now widespread rather than concentrated just north of the border area. Canada is similar in the diversity and wide distribution of its immigrants, but the distribution by country of origin is different (Canada draws primarily from China and India). In both countries many of the immigrants are highly skilled and well paid, but most are not.

Immigrant health issues have been a recurring theme in public and occupational health. In the late nineteenth and early twentieth century, there was much concern over the introduction of disease into the United States by recent immigrants, most of whom had poorer health status than native-born Americans and acculturated immigrants. Some of this concern was valid (although tuberculosis rates, of particular concern, were already high in the United States), though much of it was a cover for anti-immigrant politics.

Much of the U.S. economy has become dependent on lower-wage immigrant labor, including more hazardous industries such as poultry and meat packing (which has a very high rate of musculoskeletal injuries), construction, and agriculture (where risks include pesticide exposure as well as traumatic injury). Although the perception of poverty may be different relative to an immigrant's experience in the country of origin, in U.S. terms many of these workers are poor in a country with relatively high costs of living and therefore experience barriers such as poor transportation and inadequate health care. Immigrant workers in low-wage positions may be highly vulnerable because of language barriers, illiteracy (they may not be able to read

easily in their own language), low levels of education, discrimination, dependence on their employers, the rural location of many jobs and communities, and lack of political representation. They often lack the protection of workers' compensation and unemployment insurance because of their work in agriculture and in domestic work, where workers are not usually covered. Occupational health and safety training may be rudimentary or lacking entirely, and when given, it may be ineffective if not colloquially translated and provided in an accessible and culturally sensitive manner.

Agricultural workers in particular may be migrants with no settled home and may reside with their children in camps with poor sanitation and close proximity to sprayed fields and other hazards. The Migrant Clinicians Network (MCN) was established in 1984 to address the health needs of migrant (moving from place to place) and seasonal (traveling to a place to pick a particular crop) farmworkers and their families in the United States. MCN is national in scope and manages a variety of funded programs to help member physicians provide prevention and health care services.

All of these problems are much worse if the immigrant does not have legal status. Illegal, or "undocumented," immigrants lack basic protection under the law and are easily exploited because they can be turned in and forcibly repatriated at any time. A particularly egregious incident occurred in 2005, when Immigration and Customs Enforcement (ICE) tricked undocumented immigrants in North Carolina into assembling for what was billed to be a mandatory occupational health and safety training meeting sponsored by OSHA—a meeting that was then raided. Prior to the incident, OSHA had been making a considerable effort to reach out to immigrant workers, undocumented or documented, in order to raise awareness and elicit cooperation and compliance with occupational health and safety protection. There was concern that the ICE action would not only negate this effort, but would also discourage even legal immigrants, who fear harassment and job discrimination. A more detailed discussion of the plight of undocumented aliens is beyond the scope of this text, because the core problems are social rather than environmental or occupational.

## **TRAVEL MEDICINE**

Because of globalized trade, the need for travel to foreign operations by executive and technical staff, and the occasional case of an introduced disease in a returning tourist or business traveler, travel medicine has entered the mainstream of OEM. This section is only a very broad overview, because a detailed exposition of travel medicine would be vast and well beyond the scope of this book. Also, it is generally a mistake to rely on a book for information on travel medicine, unless it is updated very frequently. This section should therefore be considered a discussion of travel medicine as it pertains to OEM physicians; it is not an introduction to travel medicine.

Any OEM physician with responsibilities that include travel medicine must regularly consult the relevant authoritative sources for current information, without exception:

- Centers for Disease Control and Prevention: [www.cdc.gov/travel](http://www.cdc.gov/travel)
- Public Health Agency of Canada: [www.phac-aspc.gc.ca/tmp-pmv/index/html](http://www.phac-aspc.gc.ca/tmp-pmv/index/html)
- World Health Organization: [www.who.int/en](http://www.who.int/en), which allows access to the Weekly Epidemiological Record and other documents
- ProMed-Mail (sponsored by the International Society of Infectious Diseases): [www.promedmail.org](http://www.promedmail.org)

These sources provide current information on trends around the world and recommendations on appropriate immunizations.

Some generalizations can be made that narrow the scope of concern for routine and predictable travel. For example, yellow fever, although an important cause of disease, is not a common threat for business travelers, who rarely travel into endemic areas and are likely to be aware of it if they do. Likewise, high-altitude sickness, though very serious for the mountaineer, is unlikely to be a concern of the business traveler.

The role of the OEM physician is primarily in managing prevention, not diagnosis and treatment, which are best left to the infectious disease specialists. However, in managing prevention, the OEM physician does not have to do it alone: every substantial city in North America now has a travel clinic to which business travelers can be referred, and even when such a clinic is not at hand, guidance on prevention is freely available from sources such as those listed above.

The majority of business travel by North Americans is to Europe and elsewhere within North America. Most frequent business travelers are healthier and fitter than the average tourist, simply because of self-selection. Automobile injuries are the single most-common serious threat to health while traveling. Unsafe sexual behavior remains a leading cause of acquired infection among travelers, despite the well-known and ubiquitous threat of HIV/AIDS. Waterborne disease and poor sanitation are major problems not only in developing countries, but also in less-developed regions of many middle- and high-income countries as well as remote areas of all countries.

Higher-risk areas for business travel tend to be clustered as follows:

- Regions where malaria is endemic—chiefly the tropics—present a major risk to business travelers and to local residents, including employees.
- Hepatitis A is the most common vaccine-preventable illness among travelers; immunization should be routine for business travelers.
- Hepatitis B remains a serious and potentially life-threatening risk in much of the world and is easily prevented by immunization, which should be routine for business travelers.
- Influenza is a very common disease transmitted to travelers, who may not be protected by influenza vaccine designed for the Northern Hemisphere if they travel in the Southern Hemisphere, where strains are often different in a given year.
- Meningococcal disease is uncommon, especially outside equatorial Africa, but it is a significant risk for religious pilgrims

making the hajj to Mecca, in Saudi Arabia, whatever their socioeconomic class. Business travelers are unlikely to be at risk, but hajji (pilgrims) should be protected.

As noted earlier in this chapter, malaria is a major problem in global health and therefore travel medicine, and it is a practical problem in many industries with operations in endemic areas. Malaria is a risk in urban as well as rural areas. High frequencies of chloroquine resistance are now found widely throughout the tropics and southern Africa, including northern South America, central Africa, southern Asia, and Southeast Asia. Chloroquine sensitivity is confined largely to Central America, the west coast of Mexico, central and southern South America, the Middle East, western North Africa, and scattered locations in China.

Prevention is a major challenge for business travelers in all malaria-endemic areas. The OEM physician should follow authoritative guidelines for prophylaxis, but certain facts are noteworthy. The first line of protection against malaria is for the traveler to practice primary prevention against mosquito bites by reliably using DEET on clothing, sleeping inside permethrin-impregnated mosquito netting, wearing long-sleeved shirts and long pants, and spraying small spaces with DEET before occupying them; adherence to these measures is poor. Not taking malarial prophylaxis should not be considered an option for the business traveler: the risk is too high, especially where falciparum malaria is involved. On the other hand, the state of the art of chloroquine-resistant malarial prophylaxis is not ideal, and all viable choices involve trade-offs in terms of the risk of side effects and sufficient inconvenience to discourage adherence, which is generally poor. Mefloquine, an earlier drug of choice for prophylaxis, has become highly controversial after reports of serious behavioral side effects, and may be considered to present an unacceptable risk of liability. Resistance to doxycycline has not been reported. It is highly effective in the prevention of falciparum malaria, and it is usually readily tolerated by travelers, although it is photosensitizing; it may be considered to be underutilized. Most prophylactic drugs need to

be continued for four weeks after departure from the high-risk area, which often leads to adherence failure, but atovaquone-proguanil and primaquine require only a single week after departure, making them attractive options for short trips to high-risk areas. Standby emergency treatment (SBET) of suspected malaria is a relatively new strategy, which requires the traveler to be alert to early symptoms and to take an antimalarial drug at therapeutic doses at the first sign of fever. Recommendations vary, but the SBET strategy is designed for travelers who do not wish to take chemoprophylaxis (arguably a bad idea), who will be in endemic but low-risk areas, or who will be more than twenty-four hours away from medical care; it is not necessarily recommended.

Travelers' diarrhea without blood is an almost ubiquitous risk, even occurring at times in North America and western Europe, but it is rarely threatening to the healthy business traveler if he or she can remain hydrated. Bloody diarrhea is another problem altogether. Prevention of each involves avoiding potential fecal-oral contamination, which can be achieved by eating only well-cooked foods and consuming either heated or commercially bottled beverages. Iced beverages and salads present particular risks.

Emergency medical evacuation by air is a logistical nightmare. Commercial airlines vary in their policies otherwise, but they will never transport unstable patients. Air ambulance services and U.S. military aircraft (which charge by the hour) are extremely expensive. Emergency air evacuation only makes sense when the risks of remaining in the country are greater than the risk of making the trip, the patient is sufficiently stable to survive the trip, the treatment facilities at the destination are both superior and necessary for survival, and the disease is not readily communicable. A particular wrinkle in such services is that insurance policies and service contracts may be written either for "repatriation," which means transport to the patient's home country, or for transport to the closest facility providing adequate care, which may be the nearest adequate hospital, from where the recovering traveler must then find a way back home. There are commercial medical evacuation and advisory services, such

as International SOS, that manage such situations and other difficult situations under contract.

Routine travel by air carries a risk of deep-vein thrombosis. This problem is greater in economy class and on budget airlines, where it is more difficult to stretch out and to move one's legs. Travelers should flex their feet at the ankle periodically and move around to prevent venous stasis. Anti-thrombosis support hose pulled smoothly around the calf (never bunched to form a constrictive band, which has a tourniquet effect) reduces venous pooling, risk of thrombosis, and dependent swelling and discomfort.

Finally, jet lag—sometimes but not often called “circadian dyschronism”—is a familiar phenomenon for international travelers crossing longitudinal meridians and time zones, and there are many explanations for it. Discordance with circadian rhythm patterns (as discussed in Chapter 13) is probably the most important mechanism, but mild hypoxia, dehydration, sleep deprivation, and irregular mealtimes undoubtedly play a role. There are many folk remedies, and some progress has been made toward pharmaceutical treatment, particularly with melatonin (which seems to be effective in about half of users, as judged by a Cochrane review). The most reliable and nondisruptive measures to minimize jet lag, admittedly only partial in their effectiveness, include the following:

- Set watches to the destination time on takeoff and begin to think in terms of the time zone to be entered.
- Avoid all caffeinated and alcoholic beverages from several hours before departure to several hours after arrival.
- Eat in moderation when hungry and sleep when tired on the plane and in transit, but begin to eat meals on local time on arrival.
- Expose oneself to intense light during daylight hours on arrival, whether by going outdoors or using a commercial phototherapy device.

- Use sedatives or hypnotics for sleep sparingly if at all and only on arrival, never before or during a flight.
- Manage jet lag, whenever possible, with short naps or by scheduling time to adjust before any activity requiring mental acuity or high performance.
- When traveling westward, try to stay awake until bedtime at the destination. When traveling eastward, which is usually harder, do the best one can.

## **INTERNATIONAL AND GLOBAL ORGANIZATIONS**

The United Nations was established in 1945 in the aftermath of World War II in an effort to secure peace permanently, to protect human rights, and to make a better world. Regarding the latter, it has been remarkably successful, and although often criticized for being imperfect in regard to the first two goals, its humanitarian interventions, moral influence, and contribution to world stability have created conditions that allow progress in the human condition. The Universal Declaration of Human Rights (1948), together with revolutionary France's Declaration of the Rights of Man (1789) and the statements defining freedom in the Anglo-American tradition from the Magna Carta to the Bill of Rights (and including, most recently, the Canadian Charter of Rights and Freedoms, 1982), stand as milestones in the history of human rights and cornerstones of human rights law. Not surprisingly, within the United Nations, issues of environmental health and occupational health are perceived as human rights issues and are understood in terms of equity, ethics, and access, not unlike the emerging view of environmental justice in the United States, but with the addition that material security and protection are seen as basic human rights and not just a matter of fairness. The United Nations operates through the "UN system," a network of specialized bodies and subsidiary agencies, or "organs," which have complicated reporting and governance structures. The principal organizations for environmental health are the World

Health Organization (WHO), the United Nations Environment Programme (UNEP), and the United Nations Development Programme (UNDP, which usually takes the lead in coordinating UN activities in developing countries). Other agencies may have some role as well: the World Bank Group (which has played a critical role in environmental issues worldwide); the United Nations Children's Fund (UNICEF has been involved in children's environmental health, including water issues); the Food and Agricultural Organization (FAO); the International Atomic Energy Agency (IAEA); UN Habitat (which focuses on housing); the World Trade Organization (globalization); and others have all had important roles in environmental and occupational health. Where their wrists overlap, as in occupational health, UN agencies typically coordinate: WHO and the International Labour Organization (ILO) are the lead agencies in occupational health.

WHO is the United Nations agency for health issues and should already be familiar to any OEM physician. WHO is the single authoritative source for national and global health data. WHO ([www.who.int](http://www.who.int)) is essential for monitoring disease outbreaks worldwide, international travel recommendations, global environmental health, global environmental change as it affects health (UNEP is the lead for environmental issues in general), and global occupational health. WHO is managed by a director-general who, together with an executive board, is responsible to the World Health Assembly: a directing committee that meets every year with representation from all UN member states. Although many activities are centralized at the headquarters in Geneva, WHO has six regional offices and conducts much of its work, especially training, at the regional level. The WHO Regional Office for Europe is especially large and widespread (encompassing forty-two countries across Europe and Asia, because it includes Russia and the Asian states of the former Soviet Union) and is a leader within WHO in taking initiative and producing materials. There are many environmental initiatives. Occupational health is guided by a decade-long Global Plan of Action on Workers' Health; work in occupational health is closely coordinated with the

ILO. Much of the environmental and occupational health work sponsored by WHO is undertaken by a network of Collaborating Centers, which are centers of excellence located mostly at academic institutions; these centers conduct research, provide training, and produce literature and other tools to support WHO initiatives. These Collaborating Centers do not receive regular funding from WHO (although some expenses may be reimbursed), but the status is coveted as recognition of their contribution on a global scale.

The Pan American Health Organization (PAHO) serves the Western Hemisphere, having been founded in 1902 as the Pan American Sanitary Conference (later “Bureau”), with an initial emphasis on control of yellow fever. (As the construction of the Panama Canal illustrated, yellow fever was recognized as profoundly an occupational hazard.) The creation of WHO in 1946, a year after the United Nations itself, centralized world health initiatives as WHO absorbed preexisting international agencies. PAHO retained its autonomy, however, and in 1949 entered into an agreement by which it would serve as WHO’s regional office for the Americas, but would retain its preexisting structure, which is similar to WHO, and identity. PAHO also has programs in environmental health and occupational health.

The European Union (EU) currently (as of 2009) contains thirty-two member states and is mentioned many times in this book. Because of the size of the single, common European market and its purchasing power, EU standards have become the new *de facto* regulatory standards for world trade. Countries outside the EU, particularly Canada, have largely adopted EU standards in order to ensure access to the European market for their exports, knowing that those standards generally exceed those in other countries—including the counterpart American standards, which are increasingly dated—and are consistent with “greening” the economy over time. Prior to this, the EU had positioned itself as the world leader in standards setting in environmental and occupational health by harmonizing internal standards among its member states, a contentious but often very boring—and much mocked—process but one that has left it with a comprehensive body of regulatory documentation, reporting requirements, and legal

precedents for other countries to draw on, as they once drew on ACGIH for comparable documentation. The result is that world standards are being harmonized upward (in a positive sense). The EU has its own network of organizations in the field, including the European Environmental Agency and the EU-OSHA (the European Agency for Health and Safety at Work), among others.

Elsewhere, as in the Association of Southeast Asian Nations, environmental health and certainly occupational health are treated as national issues.

There are so many nongovernmental organizations (NGOs) dedicated to environmental health that it is impractical to list them. It is much more useful for the OEM physician to choose the area of interest and then to search for the relevant organization on the Internet.

There is only one global organization for occupational health, but it is very active. The International Commission on Occupational Health (ICOH) was founded in 1906 as the Permanent Commission on Occupational Health, in order to create a sustainable organization to carry on the work of a major international conference held that year on occupational injuries and illnesses. ICOH hosts thirty-six scientific committees, through which it does most of its work disseminating scientific research, providing training, and providing assistance, especially in developing countries. ICOH itself convenes the World (previously “International”) Congress on Occupational Health every three years, and many of the scientific committees hold important conferences on occupational health topics, some of them annually on off years from the congress. At the national level, ICOH has a network of national secretaries and is establishing organizations for some regions, such as the Asian region. ICOH’s Code of Ethics (discussed in detail in Chapter 25) is the global standard of practice and is acknowledged as such in the laws of some countries. The practical working language of ICOH is English, although it has also been bilingual in French, but key documents are available in many languages. ICOH is recognized as an NGO by the United Nations and has a close working relationship with WHO, ILO, and other global organizations with interests in the field, particularly the International

Social Security Association and the International Occupational Hygiene Association. ICOH is headquartered in Rome and can be reached at [icoh@ispesl.it](mailto:icoh@ispesl.it) or [www.icohweb.net](http://www.icohweb.net). Membership is open to all professionals in occupational safety and health, and the application process requires the endorsement of three current members.

## **RESOURCES**

- Amick BC III, Levine S, Tarlov AR, Walsh DC, eds. *Society and Health*. New York: Oxford University Press; 1995.
- Bim A-E, Pillay Y, Holz TH. *Textbook of International Health: Global Health in a Dynamic World*. London: Oxford University Press; 2009. [This is the standard text in global health and is most useful to the OEM physician as general background.]
- Humphries J. Child labor: lessons from the historical experience of today's industrial economies. *World Bank Econ Rev*. 2003;17(3):175–196.
- Keystone JS, Kozarsky PE, Freedman DO, Northdurft HD, Connor BA. *Travel Medicine*. 2nd ed. St Louis, MO: Mosby Elsevier; 2008.
- [Current statistics and guidance on the current policy of governments, UN agencies, and most major organizations in the field change rapidly and can be obtained only on the Internet. This is particularly true during the current economic recession. The OEM physician is encouraged to access authoritative Web sites for any factual information.]

## **NOTEWORTHY READINGS**

- Buttel FH, McMichael P, eds. *New Directions in the Sociology of Global Development: Research in Rural Sociology and Development*. Vol 11. New York: Elsevier; 2005.
- Caplan AL, McCartney JJ, Sisti DA, eds. *Health, Illness, and Disease: Concepts in Medicine*. Washington, DC: Georgetown University Press; 2004.
- Feachem RG, Kjellström T, Murray CJL, Over M, Phillips MA. *Health of Adults in the Developing World*. New York: World Bank / Oxford University Press; 1992.
- Guidotti TL. Development, health and the environment: one model. *OSH & Development* (Swedish Association for Occupational and Environmental Health and Development). May 2002;4:43–54.
- Guidotti TL. Occupational health and economic development. In: *Basic Occupational Health*. New York: Oxford University Press; 2010 (projected).
- Guidotti TL, Goldsmith DF. Occupational medicine in a developing society: a case study of Venezuela. *J Occup Med*. 1980; 22 (1):30–34.

- Keating DP, Hertzman C, eds. *Developmental Health and the Wealth of Nations: Social, Biological, and Educational Dynamics*. New York: Guilford; 1999.
- Lane RE. *The Loss of Happiness in Market Economies*. New Haven, CT: Yale University Press; 2000.
- Marmot M, Wilkinson RG. *Social Determinants of Health*. Oxford: Oxford University Press; 1999.
- Roberts MJ, Tybout JR, eds. *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*. New York: World Bank / Oxford University Press; 1996.
- Rodriguez-García R, Goldman A. *The Health-Development Link*. Washington, DC: Pan American Health Organization; 1994.
- Rotberg RI, ed. *Health and Disease in Human History*. Cambridge: MIT Press; 2000.
- Xue SZ, Liang YX. *Occupational Health in Industrialization and Modernization*. Shanghai: Shanghai Medical University Press; 1988. [This is an outstanding, now scarce, reference on China's early experience with modernization. Shanghai Medical University is now Fudan University.]

*This page intentionally left blank*

# CONCLUSION

## TO THE NEXT GENERATION OF OEM PHYSICIANS

The new physician in occupational and environmental medicine (OEM) has an opportunity to shape the future of field and through the field to shape occupational health and environmental health and integrity for the next generation. This conclusion is therefore written as a charge for the new OEM physician and the next generation. It is an opinionated, personal message from an author who is in sight of the end of his career to physicians who are just starting down this rewarding but unusual path. Unlike the rest of the book, it uses the first person plural, partly to cast what is said in the author's "authoritative" voice, but also to emphasize that many leaders in OEM share these views, at least individually, publicly or privately.

Occupational and environmental medicine (OEM) is unique in medicine because the field looks outward to society and technology for change and not inward to the next new drug or biomedical breakthrough. OEM is practiced in the world, not in the hospital, so it is a paradigm of the ambulatory, client- (if not necessarily patient-) centered style of medicine that the profession now strives for. OEM, or at least the occupational pole of it, has gone as far or farther than any other medical specialty in defining evidence-based guidelines. The environmental pole presents an opportunity to apply medical knowledge in new ways in and out of the clinic.

## **Conclusion**

With all this potential, OEM can be a model for the medicine of the future. We are focused on the individual in the office or clinic, but we can deal with population health and the big picture. Our field keeps people out of the hospital and gives them the opportunity for better lives. We apply the knowledge of medicine outside medical practice, in different and innovative ways. Our training model emphasizes experience in the community and with people as they live, not as they are come to us in the hospital, feeling at their worst, often helpless and separated from their real lives. Seen this way, OEM is a model for the medicine of the future—proactive, individualized, and evidence-based—rather than a relic of the past.

Fundamentally, OEM may have its ups and downs, but it will always be needed because there will always be a role for the physician in managing aspects of health and work and the environment will always have an influence on health. OEM matters to people because we all live in the environment and we all either work or depend on someone who does.

## **A Paean to Occupational Medicine**

Imagine that the scope of medicine suddenly doubled, creating new practice possibilities, ways of applying medical knowledge outside of the clinic, and an abundance of challenging cases. Imagine every day in clinic being different. Imagine the impact on medical practice of a demographic shift that suddenly brought large numbers of young adults back into primary care and that supported older citizens and encouraged them to be vigorous and productive. Imagine the problems of geriatrics given a new dimension in which life experience is considered an asset and retirement is optional, not forced by disability. Imagine new means of prevention opening up to public health and preventive medicine as well as new opportunities for the physician to practice prevention. Imagine public health and medicine joined together in a partnership where prevention and treatment go hand in hand. Imagine the vision of Kerr White, who advocated a union of medicine and public health, come to life.

Occupational medicine does all this for medicine and as much again for public health. As a specialty in practice, it is typically visible to primary care practitioners only in fleeting glimpses, yet it affects most patients from young adulthood and adolescence to well beyond retirement age. As an academic specialty of medicine it overlaps much of general medicine yet incorporates its own unique content and principles, creating an opportunity to apply old knowledge in new ways. It is the field of medicine most directly in touch with changes in technology and the economy, and together with preventive medicine and family practice it is arguably the field most in touch with society generally. It is certainly the least inwardly looking and “ivory tower” because it is practiced entirely in the “real world.” It is essentially defined by its social dimensions and draws its knowledge base from the study of populations and communities, yet it is highly technical and may be heavily clinical in some practice settings. It is a type of medical practice that is distinct in its challenges but includes elements that most physicians practice every day without realizing it. It is a medical specialty in form and content but with so few specialists in practice that other physicians for years to come will handle more cases than the specialists themselves. Intellectually, it is a bounty, with new issues and challenges that make every day unlike the one before.

Specialties are distinguished by the unique nature of their subject matter or practice. Most specialists are defined by organ system, such as cardiology or orthopedics; some by anatomy, such as otolaryngology; some by age group, such as pediatrics; and some by technique or mechanics of practice, such as anesthesiology. A few are defined by etiology or type of pathology, such as infectious disease and oncology. Others are defined by the need for a particular type of health care, such as family practice. Occupational medicine has been predominantly a specialty of the latter type but also has similarities to the etiology-based specialties in its emphasis on toxic and physical exposures.

As societies mature and become richer, they grow more averse to risk. The question is whether members of society who avoid a risk

## **Conclusion**

to themselves have a right or a privilege to impose it on others. Fundamentally, occupational medicine is about how we value our lives, how we treat those who create wealth and support us, and the risks we wish to take with our lives.

### **A Paean to Environmental Medicine**

The overarching mission of environmental medicine is to create a world fit to live and to raise children in, one in which all people are free to reach their biological and personal potential. The OEM physician helps to achieve this by identifying, understanding, and then correcting preventable external threats, hazards, and risk factors to health.

OEM is primarily defined by the study of tangible, material influences on the body that come from the outside world and the body's responses to these influences. The term "influences" is used here rather than "exposures" because it is a more general term and sidesteps the implication that the only relevant influences are chemical exposures or forms of energy. Some influences are personal and interactional, such as psychogenic stress, but they are just as relevant. However, a large class of influences does not fit this model. These are the effects of ecosystem disruption that may, for example, result in favorable conditions for disease transmission or lead to secondary effects mediated by the economy or social disruption. The influence of cultural change affects health through lifestyle and attitudes toward health-related behaviors. The influence of social equity and civil society makes it safe to create and to be more open with others. The influence of environmental quality and positive determinants of health gives people a reason to sustain rather than exploit communities, confident that their needs will be met and that the communities can become or remain good places to live.

One of the great unrealized opportunities in environmental health is to address the challenges of the built environment. We fuss on the edges with building-related problems and ignore the many subtle ways that urbanization and land use affect our health and lock us into

energy, transportation, and residential patterns that perpetuate health risks. Cities are artificial ecosystems, built by people but preserving remnants of natural ecosystems as part of their structure. Sustainability is just as important for these urban ecosystems as for natural ecosystems. A sustainable urban ecosystem, conceptually, has a lesser impact on the environment (footprint) as a whole than does a conventional city; it is a more pleasant and creative place to live, so that residents are motivated to preserve its features rather than overbuild; and it is an efficient structure that concentrates human impact and spares the negative consequences on less dense parts of its region. Above all, it is healthy and attractive, and its citizens sustain it because they want to and because of deeply held values, not because they are fearful. If the people who live in a place are not healthy and their aspirations are not fulfilled, why would they want to continue to live there or try to sustain it? This approach to enhancing the urban environment and making it compatible with, rather than antagonistic to, environmental protection is consistent with the principles of sustainable development and with environmental medicine. It is a fundamental issue in further progress and an unresolved problem in urban and regional planning. It connects the environment with help at a deeper level. This is an old idea, by the way. It was the view of the great public health advocate Max Joseph von Pettenkofer in the nineteenth century.

The environment or global ecosystem (which some more mystical environmentalists personify as Gaia, for those who need to attach a name or a face to big ideas) has value in itself apart from its utility to human beings. However, most people are concrete, not abstract. They are concerned most with how the environment affects them and their family and close others. Protecting the environment begins with protecting people so that people will see their interest in protecting the environment.

If people are harmed by what they encounter in the world of daily life or work, we all lose. If people are affected adversely, even a little, by environmental exposures, then we fail to reach our full potential as a community, as a society, or as individuals. If we do not invest in

## **Conclusion**

our environment, our lives will be the poorer. The question, therefore, is how much we value our lives. If we do not place a value on our lives, we will not invest in the environment.

As societies mature and become richer, we know that they grow more averse to risk. Members of society may choose to avoid risks for themselves by choosing where they live and insulating themselves with the protection they can afford, but the question is whether we would be better off lowering the level of risk for all of us. Fundamentally, environmental medicine, and all of environmental health, is about how we value our lives, how responsible we choose to be, and how we wish to live.

## **OEM AND ITS DISCONTENTS**

OEM at its best also represents a side of medicine that is the opposite of treating disease: the preservation of wellness. For all the time that medical students spend studying sick people, they are rarely asked to consider the objective of maintaining health. There is little place for the healthy in medicine, it would appear. The public and the health care system do not demand its inclusion in the medical curriculum.

For physicians who care about providing care to keep people healthy, rather than only treating them for illness, OEM is one way, and a very good way, to do this. However getting started in the field and ultimately reaching this potential is not so easy.

### **Getting Started**

As explained in Chapters 1 and 12, few physicians practice mainstream environmental medicine exclusively. The field merges into occupational medicine, and it is in the latter pole of the combined field that the OEM physician usually supports the practice. Thus, for all but a few practitioners, plus those who work in government and the few doing academic research on environmental topics alone, the key to a successful career is to become very good at occupational medicine and to leverage one's interests in environmental medicine.

Certain steps should be taken by every physician seeking to enter OEM practice:

1. Become informed about the field.
2. Define objectives.
3. Review the profile of local industry.
4. Become knowledgeable regarding local occupational health problems.
5. Plan the proposed service.
6. Identify staff, facilities, and resources.

Ideally, these steps should be taken in the order given. However, the sequence actually taken is usually the reverse. Hospitals, clinics, and groups tend to use whatever equipment and whichever staff are available to start a service (usually poorly planned), and they end up learning through their mistakes what the community does *not* want. The aspiring OEM physician often has to learn to make do, innovate, persuade, and inveigle to create a service he or she can be proud of.

It is often wise in the beginning to target a specific industry, employer, or common occupation, such as firefighters or police or health care workers, and to become known as the local expert. Department of Transportation commercial driver certification (see Chapter 18) has been the mainstay of satisfying careers. This is a surer path than becoming the expert on a particular hazard or clinical condition, which can be very limiting.

A good way to start in occupational medicine is to assume responsibility for the employee health service of a medium-size hospital. However, the alert OEM physician will quickly recognize that this can be very limiting. Hospitals tend to be very conservative employers. However, they are good training grounds. Directing a hospital employee health service is a good way to observe both the good and the bad in occupational medicine practice. It also provides a “captive” patient base on which to build for later expansion of services. The major disadvantage of getting started in hospital employee health is

## **Conclusion**

that “familiarity breeds contempt.” The occupational physician for a hospital is often a lone voice.

For an occupational health service to be responsive to local employers, the OEM physician must understand the hazards in local industries, the major employers in the area, the organization and ownership of local industries and their suppliers, and the special features of the work force. The economic base of the community holds many surprises. The presence of one or two major factories with obvious or unusual hazards (perhaps confined to a few workers) may blind one to the reality that most workers in those factories are exposed to more conventional hazards, such as noise. In many industries, such as the oil and gas and the chemical industries, the number of workers is very small compared to the volume of production. Many workers in any local economy are employed in automotive repair and servicing, fast food operations, agriculture, tourism (including hotels and motels), and even healthcare, usually outnumbering those in the “dominant” industry of the community. Each of these local industries generates sufficient injuries and illnesses, workers’ compensation issues, and fitness-for-duty problems to require medical services on a frequent basis. Because of their small scale, these industries also need help and have trouble getting what they need from consultants at a price they can afford.

## **New Thinking**

There are so many areas in which advances can be made in OEM. This is truly a field in which an individual physician can make a difference for an entire community, can discover a “new” disease (often entirely new, if associated with a new process or chemical), can introduce new ideas, and can successfully challenge authority (although perhaps not in the world of workers’ compensation—but that is another story).

The field of OEM needs to reexamine and rethink issues of professional training, the relationship between OEM physicians and departments of human resources, the balance between prevention

and case management, teamwork and competition among occupational health professions, cross-training in professional competence, and especially the balance between occupational and environmental medicine. The author of these volumes served as president of the American College of Occupational and Environmental Medicine (ACOEM), which has made a start on these issues. But the natural tendency of the OEM physician, guarding a small specialty, is to patch things up at the top rather than try to fix the foundation, which is what is needed.

Little effort has been made in the United States to study the experience of other countries with similar problems that have pursued different paths. France, Canada, Germany, Sweden, and Finland, in particular, have adopted different models for occupational health and hygiene services and have experienced their own transitions and troubles (especially Sweden). Some of their systems, such as physician training, work better than the American version and should be studied as adaptable models. Others may not work as well, such as the French preoccupation with periodic health surveillance, but we stand to learn from the experience of others. Certainly there is much in the performance of Britain's Health and Safety Executive that could be of value to OSHA.

### **Welcoming the New OEM Physician**

The field remains wide open to motivated physicians who wish to enter the practice in a credible manner and deliver service of high quality. Most occupational health care will continue to be provided by primary care physicians in the community rather than by formally trained specialists. There are now and will for many years be many opportunities for these physicians to develop their skills in occupational health care conscientiously and to incorporate occupational medicine into their practices as an integral and challenging part of primary health care.

However, organized occupational medicine (except for ACOEM) has not always been welcoming to these physicians. Some specialists

## **Conclusion**

in occupational medicine (to use the formal name of the specialty) see them as a threat. However, they are keeping OEM alive. Physicians who enter the field in mid-career bring life experience, medical and surgical skills, and knowledge of the community that helps the field succeed and stay relevant. In many places that do not have specialists, these physicians represent the standard of practice and uphold it perfectly well; most of them are already specialists in other fields. These physicians also expand the ranks of providers to meet the demand that specialists never can and to be a presence in places where there are no qualified specialists and probably never will be. Without them, demand would dry up for lack of supply, with employers turning to anyone with an MD and neglecting credentials and special expertise entirely.

There will always be a small number of young physicians who wish to train for highly specialized and academic careers. They will continue to be attracted by the intellectual richness and challenge of the field. Because the field has such a low profile in medical school and in the community, however, there will never be many of them. It is highly unlikely that the few physicians who have an interest in this field will be diverted from specialty training simply because they do not need it to practice occupational medicine. Their motivation and the achievement of their goals require the challenge and the prestige of a recognized specialty. They are the future leaders and they know it.

What visibility the field of occupational and environmental medicine does enjoy in medical school and to the public is largely due to those same physicians practicing in the field who entered in mid-career. Rather than treat this as a problem of intrusion on the specialty from the outside, diffusion into the field in mid-career should be considered an opportunity to expand the ranks of active, engaged, and experienced practitioners. The majority of practitioners who enter occupational medicine do so without taking residencies, for very good and practical reasons. It is currently nearly impossible for most physicians who enter the field in mid-career to earn a credential that would attest to their competence. This is a formula for los-

ing all control over quality assurance and continuing the marginalization of the practice of occupational medicine.

### ***Toward a Renewed Specialty***

Key to addressing this problem is establishing appropriate credentials. At present, physicians who enter the field in mid-career do not have an opportunity to demonstrate that their preparation is sound and to prove their practice skills in the field. Until recently, the discussion in this area has been about whether to allow or restore some lateral pathway to board certification, a move that would be blocked as a bad precedent by the American Board of Medical Specialties anyway. But why should specialists who train in a residency hold the same credential as physicians who enter in mid-career, and who are usually interested more in practice than in a specialty-based career? Maybe the pathways need to diverge.

There needs to be fresh air brought to this stale debate. There is considerable merit in a two-tier system, in which a separately titled, meaningful, rigorous, examination-based credential is made available for those physicians who enter the field laterally. The new credential would recognize added competence, not specialty certification, at approximately the current level expected by the current American Board of Preventive Medicine (ABPM).

Another key to addressing this problem of credentials that match competence is training. Regardless of whether a two-tier system is considered, the specialty should upgrade the existing residency in occupational medicine into a fellowship, one that matches the technical complexity and breadth of content of the field, by requiring candidates for the specialist credential to have three years of medical training, for a total of five years in training. This higher level of training would also be certified at a higher level, again by the ABPM but perhaps with a new designation of “occupational and environmental medicine.” What would be the practical and funding impact of this radical change? Almost nothing. Since virtually all OEM specialists already complete at least three years of medical training before they

## **Conclusion**

enter an occupational medicine residency, nothing would actually change, except that perception would match reality. So why not?

The two tiers already combine in one system for occupational medicine. This is, in fact, the system in place right now in both Canada and the United Kingdom, and it works well in both. (Both countries have their issues, but this is not one of them.)

We need new thinking in OEM to break through the impasse in training and credentialing. We have lived too long with a system that has been not only an imperfect compromise, but a dysfunctional straitjacket that has distorted, stunted, and deformed the specialty of OEM.

## **EPIPHANIES**

Sound decisions on health protection can be made only when the science is right. But even if the science is right, it does not by any means guarantee that the policy decision will be right. If the science is wrong, however, the policy decision is always wrong, and if it happens to be right for the wrong reason, nobody who understands will believe it. That is why OEM, which is profoundly social in its consciousness and history, must be based on science. Research is the taproot of OEM and the source of its strength. However, that is not what OEM is really about, no more than a tree is just about its roots and photosynthesis. Science without a social consciousness leads to a neglect of values, which lie at the core of medicine in the form of the relief of suffering, the provision of comfort, and the injunction to do no harm.

OEM is really about values. It has to do with how we choose to live and work. It is about how we treat the planet and whether we care about people we have not met yet who share it with us. It is about those who provide us with the goods and services we need and whether we need to know their names to be concerned with their health and well-being. It is about what we think is fair and right. It is about respect for the worker, about respect for the fellow citizens of the planet, and ultimately about respect for ourselves.

# **THE ESSENTIAL LIBRARY FOR THE OCCUPATIONAL AND ENVIRONMENTAL PHYSICIAN**

Occupational and environmental medicine (OEM) physicians are advised to refer first to these sources for reference and guidance, and to the “Resources” list at the end of each chapter.

All physicians engaged in the practice of OEM are advised to subscribe to the Occupational and Environmental Medicine List (Occ-Env-Med-L), which is an indispensable means of communication for health professionals in the field. Occ-Env-Med-L reaches approximately 4,000 recipients in seventy-five countries. Content includes announcements, important new developments, discussions of professional topics, and requests for information. The list is hosted by the University of North Carolina and has been moderated since 1993 by Dr. Gary Greenberg. It can be accessed at <http://occhealthnews.net/occ-env-.htm>. The archives of the List are particularly useful for locating current commentary on a particular topic.

Other invaluable sources include the following:

American College of Occupational and Environmental Medicine (ACOEM).

*Occupational Medicine Practice Guidelines*. Chicago: American College of Occupational and Environmental Medicine. Produced as a series with sequential revision. See [www.acoem.org](http://www.acoem.org).

## The Essential Library for the Occupational and Environmental Physician

- ACOEM Health and Productivity Toolkit.* Chicago: American College of Occupational and Environmental Medicine. Produced as an online resource available by subscription. See [www.acoem.org](http://www.acoem.org).
- AMA Guides to the Evaluation of Disease and Injury Causation.* Chicago: American Medical Association; 2008.
- AMA Guides to the Evaluation of Permanent Impairment.* Chicago: American Medical Association. Updated every few years. Use either the most current edition or the edition mandated by state law (in Texas, for example, the fourth).
- Greenberg MI, Hamilton R, Phillips S, McCluskey GJ. *Occupational, Industrial, and Environmental Toxicology.* 2nd ed. St Louis: Mosby; 2003.
- Hathaway GJ, Proctor NH. *Proctor and Hughes' Chemical Hazards of the Workplace.* 5th ed. New York: Van Nostrand; 2004.
- ILO Encyclopaedia of Occupational Health.* Geneva: International Labor Organization. 5th edition in preparation.
- International Commission on Occupational Health. ICOH Code of Ethics for Occupational Health Professionals. Rome. [www.icohweb.org](http://www.icohweb.org).
- Last JM. *A Dictionary of Epidemiology.* New York: Oxford University Press; 2007.
- Lauwerys RR, Hoet P. *Industrial Chemical Exposure: Guidelines for Biological Monitoring.* Boca Raton, FL: Lewis; 2001.
- Lippman M, Cohen BS, Schlesinger RB. *Environmental Health Science: Recognition, Evaluation and Control of Chemical and Physical Health Hazards.* New York: Oxford University Press; 2003.
- Rom WN. *Environmental and Occupational Medicine.* 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2007.
- Rosenstock L, Cullen MR, Brodkin CA, Redlich CA. *Textbook of Clinical Occupational and Environmental Medicine.* Philadelphia: Elsevier; 2005.

# INDEX

Note: Page numbers followed by “t” indicate that the reference is to a table on the indicated page.

- 9/11 terrorist attacks, 1359–1360, 1451, 1458
- Absence  
categories and classification of, 1163–1165, 1162*t*  
Family and Medical Leave Act (FMLA), 1993, 1156–1158, 1163  
management of, 1159–1161  
monitoring of, 1161–1167, 1274  
overview of, 1158–1159  
unexpected, covering for, 1157
- Absorption of xenobiotics  
overview of, 66, 70, 72–74  
routes of, 71*t*, 75–77
- Accelerated silicosis, 1046, 1048–1049
- Acceptable daily intake (ADI)  
(defined), 443
- Accident (term), 357–358  
*See also* Incident
- Accident/incident theory  
overview of, 366
- Accident prevention, 1278
- Accident-prone workers, theory of, 365–366
- Accommodation  
benefits of, 972
- definition and overview of, 1184–1188  
following illness or injury recovery, 1146  
impairment assessment role in guiding, 1171–1172  
for injured workers, 947, 955–956  
*versus* medical inquiry, 1152  
physician-management discussions concerning, 1155–1156  
*See also* Public accommodation; “Reasonable accommodation” (term)
- Acid rain, 747–748
- Acid toxins, gastrointestinal tract affected by, 122
- Acinus (liver), 119
- ACOEM (American College of Occupational and Environmental Medicine)  
committees developed by, 1361  
environmental medicine competencies, establishment by, 38  
membership increase in, 37  
overview of, 9  
public health involvement and, 49  
publications of, 10, 919, 1354, 1360

- ACOEM Practice Guidelines (APG),* 919, 945, 950, 953, 988, 1015  
Acquired characteristics as susceptibility state factor, 114  
Acrolein, 119–120  
Acromioclavicular syndrome, 997  
Acrylamide, 1119  
Act of God (concept), 361, 403  
Act of man (concept), 364, 403  
Action (as behavioral change stage), 496  
Activities of daily living (ADLs)  
capacity *versus* requirements for, 1176  
definition and overview of, 1179–1180  
functional requirements of, 1172  
medical or prosthetic devices required for, 1186–1187  
performance, disability impact on, 1175  
Acute care  
cycle of, 941–947, 940*t*  
downsizing accompanied by trend toward, 37  
in occupational medicine, 1250, 1279  
Acute disease epidemiology, 202–203, 204  
Acute exposure guideline levels (AEGLs), 454–456, 456*t*  
Acute injuries, industries producing, 1280  
Acute injury care, 1264, 1265–1266  
Acute low-back pain  
clinical management of, 977–979  
injury classification of, 951  
musculoskeletal disorders, other compared to, 970, 971  
*See also* Acute low-back pain;  
Chronic low-back pain; Low-back pain
- Acute myelogenous leukemia, 94  
Acute pain, 930–931, 1001  
Acute radiation syndrome, 528–529  
Acute renal failure, 116, 118  
Acute silicosis, 1047–1049  
ADA (Americans with Disabilities Act)  
accommodation definition and requirements drawn up by, 1184–1185, 1187, 1188  
disability and impairment concepts as defined by, 1173–1174, 1175, 1176  
fitness for duty concept impacted by, 1136–1137  
functional capacity evaluation and, 1168, 1169  
impact and overview of, 1132–1134  
medical confidentiality protected under, 1525  
productivity research and, 1240  
workers with health risks and disabilities protected through, 237  
Addams, Jane, 27, 1204  
Addictive behavior, recovering and managing, 1232  
Adjustment (defined), 266  
Administration and finance in incident command system (ICS), 1377–1378  
Administrative controls as occupational hazard control strategy, 850  
Adson's maneuver (for thoracic outlet syndrome), 992*f*  
Adult respiratory distress syndrome (ARDS), 1048, 1067  
Adults, safety for, 370  
AEGLs (acute exposure guideline levels), 454–456, 456 *t*

- Aerobics, 1230  
Aerodynamic diameter (of particle)  
(defined), 129  
Aerosol-gas mixtures, 136  
Aerosols, 125, 126–133, 129f  
Aerotitis media (barotitis), 546  
Age-specific risks for diseases, 1344  
Agency for Healthcare Research and  
Quality (AHRQ), 1212, 1213  
Agency for Toxic Substances and Dis-  
ease Registry (ATSDR), 11,  
42–43  
Agent Orange, 598, 727  
Aggravation asthma (preexisting  
asthma), 1077–1078  
Aggravation *versus* exacerbation,  
1176–1177  
Aggravational causation, 1411  
Aging  
concepts of, 481–482  
as low-back pain factor, 985  
as lung cancer risk factor, 211  
preventive medicine and, 461  
Aging population, health care for, 1208  
“Agreed medical examiner” (term),  
1334  
Agrichemical testing, 188  
Air Carrier Access Act, 1135  
Air conditioning role in limiting air  
pollution exposure, 755  
Air pilots  
fitness and medical criteria for,  
1153  
safety-sensitive evaluations for, 1154  
Air pollution  
defined, 743  
dispersion of, 748–751, 748f  
exposure to, child *versus* adult, 801  
health effects associated with,  
751–755, 750t  
incidents of, 744  
inhalation toxicology and, 124  
landfill and dump site fires as source  
of, 772  
particulate matter in, 129–130 (*see*  
*also* Particulate air pollution)  
reducing, 744–745  
regulation of, 755–757  
respiratory tract affected by,  
141–142  
time series analyses of, 241  
types of, 744–748, 743t  
in U.S. metropolitan areas, 38, 744,  
744t, 751  
*See also* Indoor air pollution  
Air Pollution Study Project, 751  
Air-purifying devices, 855–857  
Air quality  
evaluation of, 719–720  
management of, 751  
overview of, 743–744  
standards and enforcement for, 1444  
*See also* Ambient air quality; Indoor  
air quality  
Air supply devices, 854–855  
Air travel, health risks of, 1575–1576  
Airborne contamination, 797  
Airline pilots, safety-sensitive evalua-  
tions for, 1154  
Airways  
anatomy of, 130  
particle behavior in, 131–132, 131f  
Alcohol  
birth defects caused by, 177  
as lung cancer risk factor,  
211, 226  
respiratory tract affected by, 142  
smoking associated with, 226  
testing for, 1190, 1191  
toxicity from, 62–63

- Alcohol abuse  
absence related to, 1165  
identifying, 1167  
overview of, 570  
program addressing, 1233  
workplace impact of, 1189
- Alcohol compounds  
definition and overview of, 569  
hydrocarbons compared to, 671  
types of, 570–571
- Alcohol dehydrogenase elimination, 93–94
- Alcohol metabolizing systems, 93–94
- Aldicarb, 649
- Alice in Wonderland* (Carroll), 1114
- Alkali toxins, gastrointestinal tract affected by, 122
- Alkylating agents as carcinogens, 152
- Allergic contact dermatitis (eczema), 1024–1025, 1025/*f*
- Allergies, building-associated triggers of, 700
- Allergy treatment as environmental medicine, 40
- Allostasis (concept), 811
- Allyl alcohol, 119–120
- Alternative medicine, 40, 41, 45–46
- Altitude and pressure, 544–550
- Aluminum, potential toxicity of, 572–573
- Alveolar injury, diffuse, toxic gases associated with, 1064–1066, 1062/*t*
- Alveolar macrophage, 140, 141
- Alveoli, particle deposition in, 132–133
- Alzheimer's disease, 572–573
- AMA Guides to the Evaluation of Permanent Impairment* (*AMA Guides*), 1179, 1180, 1181, 1182, 1183
- Ambient air quality, 713
- Ambient environment, 3–4, 5
- Amebiasis, 763
- American Academy of Disability Evaluating Physicians (AADEP), 1183
- American Academy of Environmental Medicine, 937
- American Academy of Pediatrics, 48
- American Association of Occupational Health Nurses, 1253
- American Board of Independent Medical Examiners (ABIME), 1183
- American Board of Industrial Hygiene, 1254
- American Board of Medical Specialties, 8, 38
- American Board of Preventive Medicine, 8, 9
- environmental medicine competencies, policies concerning, 38, 39, 741
- purpose of, 459
- American College of Occupational and Environmental Medicine. *See ACOEM (American College of Occupational and Environmental Medicine)*
- American Conference of Governmental Industrial Hygienists, 135, 195–196
- American Industrial Hygiene Association, 876, 1254
- American Journal of Health Promotion*, 1221
- American Medical Association, 29, 30, 31
- American Public Health Association (APHA), 30–31
- American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 696–697, 698, 715

- Americans with Disabilities Act. *See* ADA (Americans with Disabilities Act)
- accommodation definition and requirements drawn up by, 1184–1185, 1187, 1188
- disability and impairment concepts as defined by, 1173–1174, 1175, 1176
- fitness for duty concept impacted by, 1136–1137
- functional capacity evaluation and, 1168, 1169
- impact and overview of, 1132–1134
- medical confidentiality protected under, 1525
- productivity research and, 1240
- workers with health risks and disabilities protected through, 237
- Ames, Bruce, 154, 167–171, 193
- Ammonia, 574
- Amphipoles, 575, 578
- Analgesic dependence, 1005
- Analogy, reasoning by, 215
- Anatomic loss, 1180
- Anemia, 616–617, 619–620
- Angiosarcoma, 121
- Anglo American (firm), 1563
- Animal-borne diseases, 1124, 1126
- Animal studies
- alternatives to, 188, 189
  - aluminum used in, 573
  - animal-to-human extrapolation, difficulties of, 118, 170
  - attitudes concerning, 188
  - carcinogenesis, 151, 157, 166, 167–169, 170, 171, 181, 191–192, 193, 678, 680
  - criteria for, 188–189
  - in developmental toxicology, 178
  - experimental design, 182–194
- glycol ethers used in, 571–572
- in inhalation toxicology, 144, 145–148, 146f
- reproductive and developmental hazards, 1103
- in stress response, 814, 815
- in toxicology, 58, 118, 181, 182–194
- Annual leave, 1157
- Annual physical, 1212, 1214–1216
- Anomie (concept), 810
- Anonymous data in exposure records, 1285
- Antabuse, 93
- Anthrax, 1124, 1360
- Antibiotic chemotherapy, 486
- Anticholinergics, respiratory tract affected by, 142
- Antifreeze agents, 570, 571
- Antioxidant systems, 91–93, 92f
- Architectural Barriers Act, 1134
- Archives of Environmental and Occupational Health*, 11
- Area monitoring, 272, 323–324
- Area sampling, 320
- Area under the curve (AUC) in elimination, 100–101
- Armored gloves or gauntlets, 872
- Arousal-performance curve, 812, 810f
- Arsenic
- carcinogenic properties of, 153, 1091
  - health effects of, 1114
  - human and animal effects compared, 170
  - intake, determining, 681
  - reproductive problems caused by, 174
  - technologies associated with, 1016
  - water contamination with, 764
- Arsine toxicity, 118, 1112

- Arthritis, 1237  
“As low as reasonably achievable” (ALARA), 434  
“As low as reasonably practical” (ALARP), 434–435  
Asbestos  
bans on, 581  
bodies, appearance of, 1057–1058, 1057*f*  
carcinogenic properties of, 165, 231, 1096–1097, 1098–1103, 1098*f*, 1100*f*  
definition and overview of, 574–575, 578  
exposure risks, 35  
historic uses of, 574–575, 580–581, 582–584, 581*t*  
manufacturers, legal action against, 1393, 1394  
non-fibrous counterparts of, 580  
nonmalignant disorders related to, 1048–1058, 1050*t*–1051*t*, 1054*f*  
occupational hazard and exposure effects, 150, 162, 574, 575, 580–582, 582–587, 581*t*, 716, 1403, 1405–1406  
removal of, 884  
types of, 575–578, 574*t*–575*t*  
Asbestos Hazard Emergency Response Act (AHERA), 127, 586  
Asbestos-related disease, workers’ compensation claims for, 1336, 1340  
Asbestosis  
appearance of, 1052*f*  
causes of, 575, 582  
overview of, 1051, 1054–1055  
Asphyxiants, 1110–1111  
Assaults, 1357, 1362, 1379  
Assistive devices in workplace, 1185  
Association  
apparent, chance as cause of, 216  
consistency in, 227  
despite exposure-response relationship absence, 214  
detecting, 220  
errors in, 223–224  
experimental, 215  
exposure–outcome association, 228  
factors associated with both measurement and outcome, 226  
false, 285  
group *versus* individual, establishing, 1407  
strength of, 212–213  
types of, 230  
Association of Occupational and Environmental Clinics, 1020, 1264  
“Assumption of risk” (legal concept), 1324  
Asthma  
aggravation of existing, 1015, 1017–1018  
air conditioning benefits to persons with, 755  
air pollution effect on, 748, 753, 755  
building-associated triggers of, 700, 701–702  
definition and overview of, 1068–1069  
exacerbation *versus* aggravation of, 1176–1177  
occupational (*see* Occupational asthma)  
“At permanence” (term), 1176  
Ataxia-telangiectasia, 163  
Atomic Energy Act, 1954, 1472

- Atopy, 701–703  
Attribution, apportionment compared to, 1342  
Audiograms, 1088, 1089  
Audiologists, 1254–1255  
Audit, 1561  
Autism, 256  
Automated industries, worker injuries in, 1279–1280  
Automobile injuries, 1572  
Autonomic nervous system, 61  
Avian influenza, 1127  
Aviation Medical Examiner program, 1154  
Aviation Medical Examiners, 1154  
Azide, 1111
- Back, anatomy and mechanics of, 979–980  
Back injuries, prevalence of, 951, 958  
Back pain  
  prevention program, elements of, 988*t*  
  productivity costs and expenses of, 1237  
  in service industries, 1280  
  workers' compensation claims for, 1335, 1336  
Baetjer, Anna, 28  
Barbiturates, respiratory tract affected by, 142  
Barrier creams (for hands), 867–868  
Basic occupational and environmental history, 895  
Basic occupational health services (BOHS), 1552  
Basic *versus* applied science, 1482–1483  
Bauer, George (Georgius Agricola), 14, 63
- Behavioral change, 489–501, 494–497, 498*t*–499*t*  
Behavioral controls, 850–851  
Behavioral medicine, 1206  
Behavioral toxicology, 183  
Bel scale, 540–543, 539*t*  
Bell, Alexander Graham, 540  
Benchmark, function of, 263, 264–265  
Bendectin, 180  
Benefit-cost analysis in risk management, 437–438  
Benign pleural effusion (pleuritis), 1056–1057  
Benzene  
  biotransformation of, 94–96, 95*f*  
  carcinogenic properties of, 165, 166–167  
  overview and health effects of, 676–677  
Benzo(a)pyrene  
  covalent binding to DNA, 155  
  nucleophilic "bay region" of, 153*f*  
Bereavement, absence related to, 1164  
Beryllium disease, 118  
"Best available control technology" (BACT), 434  
Beta-blockers, 222  
Beyond reasonable doubt (concept), 433  
Bhopal, India, chemical disaster in, 774, 779, 1379  
Bias  
  association, false resulting from, 285  
  of community residents and activists, 240  
  consistency mimicked by, 227  
  defined, 224  
  in health evaluations, 278  
  identifying, 224–225, 230

- Bias (*continued*)  
occupational studies prone to, 235–238  
ruling out, 228
- Bicipital tendonitis, 996
- Bile ducts, 121
- Binding agents, 97
- Binding of xenobiotics, 79–80
- Binding proteins, 79, 82
- Binge drinkers, absenteeism among, 1162
- Bio-accumulation, 798
- Bioactivated toxic agents, liver vulnerability to, 119
- Bioaerosols, 128
- Bioassays, 191–192
- Bioavailability, 60, 68–69
- Bioindicator technology, 1481–1482
- Biological agents, 1386–1387
- Biological exposure indices (BEIs), 195–196, 198  
acceptable, requirements for, 683  
defined, 454  
determining/measuring, 456 (policies for), 684–685  
pros and cons of, 455
- Biological exposure measurements, 308–309
- Biological plausibility, 214
- Biological variation (defined), 114
- Biological warfare, 128
- Biologically productive environment, 44
- Bio-magnification, 798
- Biomarkers  
of disease risk, 196, 197  
of effect or response, 196, 197  
of exposure, 196–197, 317, 318–319, 683  
exposure and response, relationship to, 195f
- overview and types of, 195–199, 262–263  
role of, 312f  
subclinical findings identified through, 260  
of susceptibility, 196
- Biomedical research, commercial applications of, 1439
- Biomedical science, environmental medicine impacted by, 1481–1482
- Biomonitoring/biological monitoring  
overview of, 310–313  
programs, 454–456  
and tracking, 194–199
- Bioterrorism, 128
- Biotransformation  
definition and overview of, 83–85, 84f  
environmental, 797, 798  
Phase II, conjugation reactions in, 89f  
in soil contamination, 797  
*See also Term* biotransformation of under specific substance, e.g.: Benzene: biotransformation of
- Bird droppings, diseases caused by exposure to, 695
- Bird fancier's lung, 1081
- Birth defects  
causes of, 176–177  
occupational and environmental exposure as factor in, 1105  
types and consequences of, 178–179, 180
- Bis-chloromethyl ether (BCME), 1095, 1096
- Bismarck, Otto von, 26–27, 1326
- Black lung, 1036–1037, 1059–1063, 1059f, 1500

- Black Lung Benefits Program, 1412, 1429
- Bladder, chemical carcinogens affecting, 152
- Bladder cancer, 1096, 1097
- Blaming the victim, incident cause theories based on, 363–364
- Bleak House* (Dickens), 1391
- Blindness, 93
- Blood, chemical carcinogens affecting, 152
- Blood-brain barrier, 80, 82–83
- Blood concentration, pharmacology and toxicology application to, 83
- Blood-retina barrier, 82
- Blood-testis barrier, 82
- Bloodborne Pathogen Standard, 1457
- Blue-collar workers, health promotion programs limited for, 1227
- Body parts, freezing of exposed, 553
- Boiling liquid expanding vapor explosions (BLEVEs), 779
- “Bona fide occupational requirement” (BFOR), 1190
- Bone, 81, 153
- Border health issues, 1564–1570
- Boundary layer, 749
- Bovine spongiform encephalopathy (BSE) (mad cow disease), 790
- Brain cancer, 222
- Breaks, frequency and duration of, 513
- Breast cancer
- detection of, 162–163
  - as occupational cancer, 154
  - risk of, 174
- Breast feeding, 1158
- Breast-feeding infants, xenobiotic exposure sources of, 81
- Breast milk, toxic agents in, 81, 99
- Breath testing, 1190, 1191
- Breslow, Lester, 1212
- BRIC countries (Brazil, Russia, India, and China), 1544
- British anti-Lewisite (BAL), 616
- British Columbia, workers’ compensation in, 1339
- British medicine, North America influenced by, 20–22
- British occupational medicine, 24, 34
- British public health movement, 23
- 1-bromopropane, 677
- Bronchiolitis obliterans, 634
- Bronchitis, 140, 755, 1084
- Bronchogenic carcinoma, 575
- Brucellosis, 1124
- Building-associated outbreaks
- building-related disease, 694–696
  - chemical exposure, evaluating in, 718t
  - definition and overview of, 689–692
  - differential diagnosis of, 689, 688t
  - epidemic curve for, 730f
  - interventions for, 734–735
  - investigating, 717–720
  - propagation trees for, 731f
  - triage of, 692–694
  - triggers, 700–703
- See also* Psychogenic outbreaks; Sick building syndrome (SBS); Tight building syndrome (TBS)
- Building Owners and Managers Association (BOMA), 734
- Built environment
- ambient environment impact on, 5
  - children’s environmental health affected by, 802–803
  - definition and overview of, 3, 4
- Burden of proof, 1418–1420

- Burnout, 820, 831–837  
Burns, occupation-related, 968–970  
Bursitis, 995–996  
Bush, George W., 1451, 1458, 1468  
Business travel, health issues and risks in, 1571–1573
- Cadmium, 116, 117  
Calhoun, John B., 814  
California, workers' compensation in, 1329–1330, 1341, 1412  
California Air Resources Board, 751  
California Occupational Health and Safety Administration (CalOSHA), 727, 728, 729, 1462–1463  
CaNa<sub>2</sub>=EDTA, 616  
Canada  
British medicine influence on, 21, 34  
British occupational medicine influence on, 34  
British public health movement influence on, 23  
drug screening in, 1189–1190  
employment law in, 1132–1133  
health promotion in, 1222  
labor conditions in, 1323  
regulatory agencies in, 1473–1476  
workers' compensation in, 1321, 1327, 1336, 1339, 1352  
Canada Pension Plan (CPP) disability benefits, 1177, 1178  
Canadian Association of Physicians for the Environment (CAPE), 48  
Canadian Aviation Medical Examiners, 1154  
Canadian Board of Occupational Medicine, 9  
Canadian Human Rights Act, 1977, 1132
- Canadian Human Rights Commission, 1190  
Canadian Institute for Advanced Research (CIAR), 487, 489  
Canadian Mortality Data Base, 290  
Canadian Society of Medical Evaluators, 1183  
Canaries in coal mines, 1499  
Cancer  
air pollution as factor in, 754  
asbestos-related, 582  
biology of, 155  
in chimney sweeps, 64, 154  
cluster investigations of, 238  
death from, 482  
descriptive studies of, 242  
detection of, 162–163  
environmental causes of, 1090–1091  
industrial activity, relation to, 254  
mortality and prevalence of, 149–150  
occupational, 1091–1098  
overview of, 1089–1090  
predisposing factors in, 163–164  
proportionate mortality (or morbidity) studies on, 249  
public perception of, 149, 150–151  
risk and risk factors, 155, 227–228, 263 (*see also under specific risk, e.g.: Radiation-induced cancer*)
- Cannon, Walter, 815  
Capital-intensive high-technology industries, worker injuries in, 1279–1280  
Caplan's syndrome, 1063  
Carbamate pesticide, 646, 648–656  
Carbaryl (Sevin®), 649  
Carbofuran, 649  
Carbon dioxide, 772  
Carbon disulfide, 1114–1115

- Carbon monoxide  
 in building environment, 698  
 in confined spaces, 1107  
 exposure effects of, 1110, 1111  
 exposure-response relationship, clinical for, 589t  
 lung not affected by, 134  
 overview and toxicity of, 588–593  
 as pollutant category, 756  
 sources of, 698
- Carcinogenesis  
 DNA damage associated with, 158–159, 158f  
 epigenetic, 164–167  
 overview of, 148–151  
 sequence of events in, 155–163, 156f  
 testing controversies, 167–171  
*See also* Chemical carcinogens; Respiratory carcinogenesis
- Carcinogenicity  
 bioassays for, 191–192  
 toxicity relationship to, 170–171
- Carcinogens  
 biotransformation of, 157  
 classification system for, 151–152  
 exposure, child *versus* adult, 801  
 human, identifying, 171  
 in workplace, 1091t–1092t
- Cardiotoxic agents, 1111
- Cardiovascular disease  
 interventions, employer role in, 1239  
 prediction of, 475–476
- Cardiovascular effects of air pollution  
 overview of, 754
- Cardiovascular fitness training, 1223–1224
- Cardiovascular risk  
 reductions in, 237
- Cardiovascular risk factors  
 modification programs, 1231–1232  
 productivity costs and expenses of, 1237  
 screening for, 1237
- Carpal tunnel syndrome  
 causes and risk factors for, 992–993  
 diagnosing, 998f  
 focus on, 992  
 overview of, 999–1001
- Carroll, Lewis, 1114
- Carson, Rachel, 660
- Case  
 cumulative count of, 113–114  
 definition and recognition of, 112
- Case management  
 overview of, 1236–1238  
 wellness and productivity management, role in, 1239
- Case-referent studies  
 as cluster investigation study design, 238  
 as community environmental health assessment study design, 240  
 overview of, 244, 250–252  
 prospective (cohort) studies compared to, 256, 271
- Case reports, 257, 258
- Case series, 257–258
- Catalase, 91–92
- Catastrophic illness, 1238
- Categorical pollution, 772
- Cauda equina syndrome, 974
- Causalgia, 1006
- Causality (term), 1402
- Causation  
 definition and overview of, 209, 1402–1412  
 epidemiological evidence for, 211–212

- Causation (*continued*)  
  etiological investigation, 209–211  
  likelihood assessment, Hill criteria  
    for, 212–216, 213*t*, 229
- Cause  
  apportioning different, pitfalls of,  
    231–232  
  defined, 229  
  ruling out, 224–228  
  types of, 230  
  workers' compensation apportion-  
    ment by, 1342–1343
- Cause–effect relationship  
  epidemiology and, 210  
  overview of, 228–232
- Cause in fact (defined), 1405
- Cell, target, carcinogen delivery to,  
  155
- Cell growth, undifferentiated, 159
- Cell membrane  
  absorption across, 73–74  
  lipid bilayer of, 73*f*
- Cellular receptors, 123
- Centers for Disease Control and Pre-  
vention (CDC)  
  children's health issues, involvement  
    in, 806  
  environmental data disseminated by,  
    774  
  influenza pandemics, potential  
    monitored by, 1381  
  productivity concerns addressed by,  
    1209  
  travel medicine information  
    provided by, 1571
- Certainty role in litigation,  
  1412–1423
- Certification of Evaluation in Disabil-  
ity and Impairment Rating  
(CEDIR), 1183
- Certified industrial hygienists (CIHs),  
  846
- Cervical syndrome, 992–993
- Chamber studies (inhalation toxicol-  
ogy), 144, 145–147, 146*f*, 148
- Chance  
  identifying, 230  
  ruling out, 216–218, 228
- Charter of Human Rights and Free-  
doms (Canada), 1132
- CHD Meridian, 1259
- Chekhov, Anton, 46–47
- Chelation, 615–616
- Chemical agents, 1386
- Chemical carcinogens, 149, 151–153,  
  153–163, 169
- Chemical contamination of water,  
  763–768
- Chemical exposure  
  in building-associated outbreaks, 718*t*  
  noise exposure compared to, 273  
  outcome, genetics influence on, 206  
  public health triad applied to,  
    473–474, 473*f*
- Chemical hazards  
  building-related diseases, role in, 711  
  emergency management and,  
    1379–1380  
  overview of, 775–777
- Chemical reactivity, 133, 134
- Chemical safety, 777–781, 1380
- Chemicals  
  body production of, 56  
  consumption, accidental of toxic,  
    75, 76  
  exposure, environmental to, 171  
  fate and disposition of, 66  
  hazard level, identifying for, 64–65  
  respiratory tract affected by, 142  
  testing of, 188

- See also* Synthetic chemicals  
Chemists, cancer risks among, 1091  
Chemoprevention (defined), 155  
Chemotherapy, 82  
Chest films, 1285  
Chevalier, Maurice, 482  
Chilblains (pernirosis), 552  
Child-care leave, 1163  
Child labor  
  among lower class, 22–23  
  in developing countries, 802,  
    1554–1557  
  historic conditions, 1500  
Childhood mortality, 485  
Children  
  air pollution exposure impact on,  
    755, 801  
  environmental health and, 801–807,  
    1450  
  hazardous substance exposure of, 781  
  lead poisoning in, 81 (2XP: lead)  
  safety for, 370  
Chimney sweeps, cancer in, 64, 154  
Chiropractic manipulation, 982  
Chivian, Eric, 45  
Chlordane, 660  
Chlorine, 760, 765  
Chloroethanes, 677–678  
Chloromethane (methyl chloride), 678  
Cholestasis, 121  
Cholestyramine, 97  
Cholinergic muscarinic receptors, 123  
Chromium, 682  
Chronic bronchitis, 140, 755  
Chronic disability, 1238  
Chronic diseases  
  epidemiology of, 202–203, 204–205  
  health, improving in persons with,  
    1201  
Chronic disorders, 202  
Chronic illness  
  absence and, 1162, 1165  
  age-related, 1208  
Chronic low-back pain  
  ability to work despite, 971  
  clinical management of, 978–979,  
    980–981  
  disease classification of, 951  
  factors affecting, 970–971  
  transition to, 974–975  
Chronic obstructive pulmonary disease (COPD)  
  air pollution effect on, 753  
  susceptibility to, 164  
Chronic pain  
  acute pain compared to, 930  
  amplification of, 931  
  overview of, 1001–1006  
  recovery from, 983–984  
  syndrome, 974  
  workers' compensation claims for,  
    1336  
Chronically ill, 755  
Chrysotile, 575, 578  
Cigarette smoke  
  as carcinogen, 1090  
  constituents of, 715*t*  
  content of, 140  
  as indoor air pollution source, 716  
  passive, exposure to, 125, 802, 1450  
  persons with atopy affected by, 702  
  sidestream cigarette smoke, 702,  
    716, 715*t*  
Cigarette smoking  
  air pollution exposure compared to,  
    142  
  asbestos exposure interaction with,  
    35, 231, 575, 584  
  asbestos exposure risks compared  
    to, 1099

- Cigarette smoking (*continued*)  
as low-back pain risk factor, 985  
lung cancer association of, 154,  
211–212, 231, 1090, 1091, 1403  
workplace exposure interaction  
with, 227–228
- Circadian rhythms, 826, 827–828,  
1575
- Cirrhosis, 121
- Cis-retinoic acid, 155
- Clean Air Act, 1970, 755, 1448, 1451
- Clean Water Act, 1976, 1448
- Cleaning agents, 716
- Clearance, 10
- Climate, health issues associated with, 3
- Clinical and laboratory tests  
precautions concerning, 279  
toxicity determination, obstacles to,  
110–111
- Clinical dose- or exposure-response  
relationship, 106, 108–111
- Clinical encounters, 890–893
- Clinical epidemiology, 207, 256
- Clinical evaluation, 889–890
- Clinical exposure-response  
relationship, 111
- Clinical medicine  
emphasis of, 889  
evidence accepted in practice of,  
1439  
preventive medicine incorporated  
into, 460, 461  
as public education vehicle, 890
- Clinical occupational medicine,  
1016–1020, 1261
- Clinical preventive medicine, 1218
- Clinical standard in risk management,  
432
- Clinical testing, 261–262
- Clinical toxicology, ingestion in, 75
- Clinical trials, randomized in occupa-  
tional epidemiology, 234, 235
- Clinton, Bill, 1457
- Cluster investigations, 238–239
- Clusters, 239, 240
- Coal mining, 1499–1501
- Coal tar products as skin cancer risk  
factor, 64
- Coal workers, health problems of, 22
- Coal workers' pneumoconiosis,  
1036–1037, 1059–1063, 1059/  
1500
- Coarse particulate matter, 756
- Cocaine, 142
- Cochrane Collaboration, 501
- Codes of ethics  
basic principles of, 1517–1518  
duties and obligations, 1518–1529  
overview of, 1514–1517
- Codex Alimentarium, 624
- Cold air-induced asthma, 1070–1071
- Cold temperature, 143
- Cold water, 554
- Cold weather, 556–558
- Coliform count, 762
- Collective bargaining agreements,  
1276
- Colon cancer risk factors, 123
- Commercial Drivers Medical Exam-  
ination (CDME), 1153–1154
- Commission for Environmental  
Cooperation of North America  
(CEC), 1478
- Communicable diseases  
employee screening for, 1145  
propagation pattern of, 243, 243/  
f
- Communications devices, 1186
- Community, 1224
- Community-acquired infection,  
1126–1128

- Community Awareness and Emergency Response (CAER) program, 430
- Community-based health care, 26
- Community-based health promotion programs, 1225
- Community-based healthcare facilities overview of, 1256–1257, 1260–1261 types of, 1257, 1259, 1261–1269
- Community-based occupational medicine practitioners, 1252, 1253
- Community environmental health assessment studies, 239–240
- Community hospitals, 1267
- Community medicine, 1217, 1220–1221
- Community right to know, 429–430
- Companion animals in workplace, 1185
- Comparison population, 264
- Compartments (of body) definition and overview of, 99–100 elimination rates for, 101–104
- Complex regional pain syndrome, 1006
- Compliance in clinical medicine barriers to, 460, 477 behavioral change, motivating and sustaining, 489
- Compliance in occupational and environmental medicine, 489–490
- Complicated silicosis, 1044–1045
- Component causes, 230
- Composting, 770, 771
- Comprehensive Environmental Response, Compensation, and Liability Act, 1980 (CERCLA), 786
- Comprehensive occupational and environmental history, 896–897
- Compression, 545–547
- Computer workstation ergonomics, 512, 513–514
- Confidentiality, patient in employee assistance programs, 1234, 1235, 1285 for executives, 1215 law and legislation, 276 medical evaluation, rules and regulations governing, 1142, 1152 occupational health records, safeguards for, 1250, 1285, 1286 occupational health services, hierarchy and, 1274 risks to, 328–329
- Confined-space incidents, 1107–1109
- Confounding phenomenon confounders defined, 285 overview of, 226–227 ruling out, 228
- Conglomerative silicosis, 1044*f*
- Conjugation in xenobiotic biotransformation, 84–85, 88, 89*f*, 90
- Consciousness, loss of, 1109–1111
- Consequence analysis tool, 880*t* defined, 876
- Consequence management (term), 380
- Consistency as causation criterion, 213–214
- Construction fatality rates in, 960 injury rates in, 952–953, 960, 961*t*
- Constructive stress, 812–813
- Consultation practice in occupational medicine overview of, 1260, 1261, 1264–1266
- Consumer health and safety, 39

- Consumer Product Safety  
Commission, 1468–1471
- Consumer product toxicology, 70, 72
- Contact dermatitis, 1024–1030, 1025f
- Contact urticaria, 1026
- Containment and isolation as occupational hazard control strategy, 849
- Contaminated clothing, handling, 866–867
- Contemplation (as behavioral change stage), 495–496
- Contractors' employees, documentation for, 282
- Control banding  
Deming cycle application to, 877  
as manager and supervisor tool, 884–885  
overview of, 885–887
- Control group  
alternatives to, 247  
definition and function of, 246
- Control of Occupational Substance Hazardous to Health (COSHH), 885
- Conversion reactions, 934–935
- “Corporate culture” programs, 1225
- Corporate medical departments  
closing of, 36, 37  
occupational medicine involvement of, 33  
overview of, 1269–1275
- Corporate medical directors  
duties and authority of, 1251, 1252–1253, 1273  
emergency management role of, 1374–1375
- Corporate occupational health services  
corporate medical department, 1269–1275
- labor-management relations, 1275–1277  
overview of, 1269–1270
- Corporate responsibility, 13
- Corporate wellness movement  
overview of, 1207
- Corporatist practice of medicine, attitudes concerning, 30
- “Correct and adapt,” risk  
management mid-course corrections through, 879–880
- Cosmetics, consumer safety and, 70
- Cost-effectiveness analysis in risk management, 437
- Coughing, function of, 137
- Council of Accreditation in Occupational Hearing Conservation, 1088, 1255
- Counselors, referral to outside, 1236
- Country (defined), 1543
- “Covered entity” (term), 1133–1134
- CPR, 1109
- Crisis intervention, 1233
- Critical appraisal, 1439
- Critical science, occupational and environmental medicine as, 1482–1497
- Crocidolite, 578
- Cross-sectional studies  
design of, 240  
ecological association studies compared to, 255  
out-migration following work-related illness as complicating factor in, 236–237  
overview of, 244, 252–254  
reference groups in, 266
- Crowding, human *versus* animal response to, 814–815
- Cryptococcosis, 695

- Cumulative impairment, rating, 1182  
Curative/ameliorative services, 1248  
Cut size (of particle) (defined), 129  
CVS (drugstores), 1258  
Cyanide  
  exposure effects of, 1111  
  overview of, 593–597  
Cyanuric acid, 117–118  
Cytochrome-linked MFO enzymes, 86–87  
Cytochromes  
  distribution and function of, 85–86  
  MFO system, interaction with, 87–88  
  
D-penicillamine, 616  
Data management and analysis, 296  
*Daubert v. Merrel-Dow*, 180, 1396–1397, 1398  
DBCP, 660  
DDT  
  bans on, 799, 1448, 1449  
  as carcinogen, 660  
  conversion into DDD, 94  
  in mosquito control, 1125  
De Briquet's syndrome (term), 932  
De minimis risk standard in risk management, 431  
*De Morbis Artificum Diatriba (Diseases of Workers)* (Ramazzini), 16–17, 18  
De Quervain's tenosynovitis  
  diagnosing, 996f  
  overview of, 995, 998–999  
Deaf persons, telecommunications services accessible to, 1135  
Death, causes of, 482–483  
Death certificates  
  as epidemiology data source, 204, 272  
  for firefighters, 289–290  
  mortality determined through, 260  
  
Decibel scale, 540–543, 539f  
Decompression, 547–549  
Deep-vein thrombosis, 1575  
DEET, 662, 1125, 1573  
Demand-control model  
  effort-reward model relationship to, 814  
  overview of, 810, 813  
Dementia, 183, 1167  
Deming, W. Edwards, 877  
Deming Cycle (Plan, Do, Study, Act, PDSA), 877–878  
Demography, occupational and environmental medicine impacted by, 1481  
Density as gas toxicological behavior factor, 133, 134  
Deontology, 1508–1509  
Department of Agriculture, 1473  
Department of Defense  
  drug screening by, 1197  
  overview of, 1472  
Department of Energy, 529–530, 1472  
Department of Transportation, 1473  
  drug screening by, 1189, 1190, 1191, 1192, 1196–1197  
  overview of, 1471  
  safety-sensitive evaluations of, 1153, 1154  
Dependency ratio  
  definition and overview of, 1208–1209  
  for selected years, 1207t  
Dependent, care for, 1163  
Depigmentation  
  solvents as cause of, 1025f  
Depositions, 1435–1436  
Depression  
  cross-sectional study application to, 253

- Depression (*continued*)  
identifying, 1167  
occupational medicine during, 29  
workers' compensation during, 1328
- Dermatoses. *See* Skin disorders
- Descriptive studies, 242–244
- Deskilling as burnout factor, 833–834
- Detroit Industrial Clinic, 1269
- Developing countries  
occupational health services in, 1553–1554, 1562
- Developmental toxicology, 172, 176–180
- Diabetes, 1239
- Diagnosis in occupational medicine, 1250
- Diagnostic occupational and environmental history, 895
- Dialysis dementia, 573
- Dibromochloropropane (DBCP), 172
- Dichloromethane, 678
- Dickens, Charles, 1391
- Dictionary of Epidemiology* (Last), 1217
- Dietary consumption, 254
- Diethylene glycol, 571
- Diethylstilbestrol (DES), 175
- Diffusion of innovation, 497, 500
- Dimercaptosuccinic acid (DMSA)  
(Succimer), 616
- Dimethylaminopropionitrile  
(DMAPN), 1119
- Dinitrogen tetraoxide, 630
- Dioxane, 165, 166
- 1,2-Dioxane, 767
- Dioxins  
molecular structure of, 596f  
as pollutant category, 772–773  
toxicity of, 597–600  
types of, 595f
- Dipyridyl herbicides, 664
- Direct assessment of exposure, 303–309
- Disability  
age-related, 1208  
benefits, 1177–1177 (*see also* Canada Pension Plan (CPP) disability benefits)  
concerns over, 481  
defined, 1134, 1170, 1172, 1174–1175  
documentation of permanent, 1333  
evaluation of, 1171–1172, 1332, 1337–1338 (*see also* Impairment assessment)  
*versus* job requirements, bridging gap between, 1184–1185  
job-specific *versus* general, 1136–1137  
living with, consequences of, 486  
low-back pain as cause of, 984  
prevention of, 460, 461, 471, 1351–1352  
psychosocial effects of, 926–927  
reduction and limiting, 486, 1014  
severity, increasing of, 1351  
worker education concerning, 1186
- Disability-adjusted life years (DALYs)  
cancer as factor limiting, 150  
definition and overview of, 486
- Disability Discrimination Act, 1995  
(United Kingdom), 1133
- Disability management  
in occupational and environmental medicine, 53  
tertiary prevention relationship to, 471
- Disability movement, 1173
- Disability (or impairment)  
evaluations, 891

- Disability-related occupational disorders, 253
- Disability rights, 1132–1135
- Disaster
- cycle of, 1366
  - definitions of, 1365–1367
  - as occupational stress type, 842
  - preparedness for, 1366, 1368, 1369
- Discovery, legal definition of, 1434
- Disease
- air pollution as contributing factor in, 750*t*
  - definitions and concepts of, 949, 1345
  - detection and intervention, early of, 462
  - management of, 1236–1238
  - prediction of, 475–476
  - prevention of, 1013, 1016, 1221–1222
  - risk assessment for, 1013
- Disease categories, assigning, 285–286
- Disease outcomes, 241
- Disease progression, prevention of, 471
- Disease rates, dietary consumption relation to, 254
- Disease reporting or prevalence, 280
- Diseases of Occupations* (Hunter), 34
- Disinfection byproducts (DBPs), 764–765
- Disinfection Byproduct Rule, 765
- Disney, medical clinics at, 1258–1259
- Disorders
- existing, aggravation of, 1015
  - monitoring and surveying course of, 333*f*
  - multiple, studies on, 253
  - slowly developing, 1347
  - stress relationship to certain, 816, 817, 818, 816*t*
- See also* Occupational diseases and disorders; specific type of disorder, e.g.: Movement disorders
- Dispute resolution
- evidence-based, 1438–1441
  - informal *versus* legal, 1392
- Distress (term), 813
- Distributed emergencies, 1381–1384
- Distribution of xenobiotics
- overview and action mechanisms of, 78–83
  - term, 66
- DNA
- carcinogen effect on, 158–159, 158*f*, 164, 165
  - damage to, 160*f*
  - toxin effect on, 87
- Doll, Richard, 154
- Domino theory of injury cause, 366–367
- Donora, Pennsylvania air pollution episode, 1948, 38, 744, 751
- Dormancy of abnormal cells, 160–161
- Dose (defined), 106, 108
- Dose-response relationship
- concept, historic development of, 60
  - definition and overview of, 381
- Doubling of risk, 1427–1428
- Downsizing trends, 36, 37
- Drainpipe odors, 712
- “Drin” pesticides, 661
- Drinking water
- chemical contamination of, 767
  - disease exposure risks of, 763
  - disinfection of, 759–760
  - lead in, 765, 766
  - sources of, 762
- Drivers, bus and truck, medical certification for, 1153–1154

- Droplets, 127–128
- Drug abuse  
birth defects caused by, 177  
developmental abnormalities caused by, 178  
disability, exclusion from classification as, 1134  
program addressing, 1233  
treatment, mandatory for, 1192  
workplace impact of, 1189, 1193
- Drug interactions, binding role in, 79–80
- Drug screening and testing  
for commercial drivers, 1154  
employee assistance programs following positive, 1232–1233  
by government agencies, 1196–1198  
impact of, 1189, 1193  
overview of, 1188–1190  
personnel involved with, 1190–1193  
technical aspects of, 1193–1196
- Drug-using employees, profile of, 1192
- Drugs, medical administration of, 70
- Drugstores, medical clinics at, 1258
- Dumps  
fires at, 772  
as waste disposal method, 770
- Dust particles  
deposition of, 130  
inhalation of, 1033
- Dust-related disease, 1034
- Dye workers, cancer risks among, 1091
- Dynamic exposure, 144
- Earplugs and earmuffs, 863–864
- Earth Day commemoration, April 22, 1970, 1448
- Ecological association studies  
applications of, 241  
overview of, 244, 254–256
- Ecological fallacy, 254–255
- Ecological illness  
term, 938
- Ecological study designs, 234
- Econometric approaches, 255–256
- Economic development  
child labor, 1554–1557  
cycle of, 1547–1551  
environmental health, 1557–1559  
globalization and trade, 1551–1552  
occupational health, 1552–1554  
overview and basics of, 1537–1547
- Economic recession  
impact of, 1542  
occupational health services in, 1259, 1270  
workers' compensation, effect on, 1330–1331, 1350, 1353
- Economy  
emerging, 1490–1497  
global nature of, 1535  
occupational and environmental medicine impacted by, 1481
- Ecosystem  
change, health issues associated with, 3, 4, 44–45  
damage as global violence, 45
- Ecosystem health  
ecotoxicology, 796–799  
global atmospheric change, 799–801  
overview of, 795–796
- EDB, 660
- Effect  
biomarkers of, 262  
measurement and magnitude of, 322–323, 323*f*

- The Effects of the Principal Arts, Trades, and Professions, and of Civic States and Habits of Living, on Health and Longevity* (Thackrah), 18
- Effort-reward model, 814
- Egypt, ancient, occupational medicine in, 14
- Elderly, air pollution exposure, limiting to, 755
- Electrical burns, 968
- Electrical fields, 513f
- Electromagnetic spectrum  
magnetic fields and, 514–518  
overview of, 514f
- Electronic medical records, 281
- Elimination  
curve, 103  
defined, 99  
kinetics of, 105  
rate of, 100f, 101–104
- Ellenbogen, Ulrich, 14, 63
- Embryotoxicity, 178
- Emergency management  
capacity, building in, 1375–1378  
chemical hazards, 1378–1380  
distributed emergencies, 1364–1365  
history of, 1358–1365  
occupational health services role in, 1365–1376  
overview of, 1357–1358  
personnel specializing in, 1369  
services and functions in, 1374–1376  
terrorism, 1357, 1384–1387
- Emergency medical evacuation by air, 1574–1575
- Emergency medicine personnel, 1369–1370
- Emergency medicine specialists, 1261, 1263
- Emergency Planning and Community Right-to-Know Act, 1986 (EPCRA), 429, 773, 777
- Emergency response as safety function, 1278
- Emissions  
controlling and monitoring, 297, 748  
escape and dispersal of, 320–322, 749–751  
volatile, 772
- Emotional distress, 1167
- Emphysema, 755
- Employee assistance (defined), 1232
- Employee assistance programs (EAPs)  
confidentiality protection in, 1234, 1235, 1285  
limitations of, 820  
mental health practitioner referrals through, 819  
overview of, 1192, 1197, 1232–1236  
work behavior problem intervention through, 836
- Employee facilities, access barrier removal to, 1186
- Employee health programs, downsizing and outsourcing impact on, 36–37
- Employer role in injured worker recovery, 944, 946
- Employment, physical  
historic practices, 1136  
*See also* Independent medical evaluation (IME); Medical evaluations
- Employment security, 1238
- Encephalopathy, 1112–1115
- Endocrine disruption, 174–176
- Energy Department, 529–530, 1472

- Energy Employees Occupational Illness Compensation Program (EEOICP), 1322
- Energy release, uncontrolled of, 505
- Engineering, 380
- Engineering controls, 849
- England, workers' compensation in, 1326
- Enterohepatic circulation, 97–98
- Enterprise risk assessment defined, 876  
and management, 877–880
- Enterprise size, injury rate correlation to, 953
- Environment divisions of, 3–5  
exposure in, 70  
as incident element, 358
- Environmental and Occupational Medicine* (Rom), 44
- Environmental contamination, 405–406
- Environmental diseases, 1015
- Environmental epidemiology causation in, 209  
challenges of, 284t  
control groups in, 247  
exposure and response, measure and category of, 216  
occupational epidemiology, relationship to, 232  
overview of, 206–208  
study design in, 238–241
- Environmental hazards as environmental health issue, 39  
occupational hazards compared to, 739–740
- Environmental health in Canada, 1473–1474  
and children, 801–807
- economic development and, 1557–1559
- international organizations for, 1576–1578, 1579
- occupational health overlapping with, 4–5
- overview of, 739–743
- physician training in, 47
- public health functions in, 739t
- scope of, 39
- subfields of, 4f
- Environmental hypersensitivity disorder, 938
- Environmental Justice Geographic Assessment Tool, 773
- Environmental medicine biomedical science impact on, 1481–1482  
competencies for, 38  
as critical science, 1485–1489  
definition and scope of, 5, 7–8  
development and status of, 6, 7  
as emerging field, 3  
history of, 12  
occupational and environmental medicine, integration within, 37, 38–39  
occupational medicine compared to, 53, 1020  
practicing, 39–42  
responsibility for, 6  
schools of thought in, 42–46, 43f
- Environmental medicine cases, 40–41
- Environmental monitoring, 297–298
- Environmental NGOs, 1543, 1558–1559
- Environmental Protection Agency (EPA)  
acute exposure guidelines developed by, 452–453, 454t

- chemical safety role of, 1380  
children's health issues, involvement in, 804, 806  
community risk data expansion by, 430  
environmental data disseminated by, 775  
Hazard Ranking System of, 788  
history of, 1447–1452  
information disseminated by, 773  
National Ambient Air Quality Standards (NAAQS) reviewed by, 757  
overview of, 1443–1447  
radiological emergency planning conducted by, 531  
risk assessment paradigm developed by, 382*f*, 384*f*  
risk management paradigm developed by, 382*f*  
Environmental Public Health Tracking, 198–199  
Environmental regulatory compliance, 39  
Environmental Response, Compensation, and Liability Act (CERCLA), 1449  
Environmental tort litigation, 1395  
Environmental toxicology, ingestion in, 70, 72  
Epicondylitides, 997–998  
Epidemic curves, 242–243, 243*f*  
Epidemiological exposure-frequency relationship, 319–320  
Epidemiological exposure-response relationship mathematical models (or representations) of, 113*f*, 114*f* overview of, 106, 112–115  
Epidemiological study design descriptive studies, 242–244 examples of, 287–292 flow diagram for, 288*f* inferential study designs, 244–259 types, time line for, 245*f*  
Epidemiology causation in, 209–216 current challenges in, 206 definition and overview of, 201–202 environmental medicine and, 44 etiological investigation in, 209–211 historic development of, 202, 205, 206 legal aspects of, 1423–1428 methodological issues and interpretation in, 216–232 prevention science based on, 459 research, 272–283 scope of, 203, 204–205 strengths and limitations of, 208–209, 210, 211 taxonomy of, 203*f* techniques and approaches, 203–204  
Epigenetics carcinogens and, 193, 194 definition and overview of, 164  
Epoxides formation of, 87, 95–96  
Equal Employment Opportunity Commission functional capacity evaluations and, 1169 health question *versus* medical inquiry distinction made by, 1149 major life activity classification determined by, 1134  
Equipment modification, 1185  
Ergonomics defined, 951

- Ergonomics (*continued*)  
efficiency benefits of, 972  
in high-technology industries, 1280  
medical model dominated by, 42  
overview of, 506–514  
pregnancy and, 1106  
safety role of, 369
- Erosions, gastrointestinal, 122
- Error  
level, acceptable in outcome studies, 263  
risk of, 218  
tolerable probability of, 217, 218
- “Errors and omissions” professional liability insurance, 1438
- Erythrocytes, 617, 619, 617/<sup>f</sup>
- Esophageal cancer, 123
- Ethanol  
effects of, 570  
excretion of, 98  
fetal alcohol syndrome association of, 178  
metabolization of, 93  
pancreas affected by, 122
- Ethics in Western tradition, 1508–1510
- Ethnically isolated workers, 1227
- Ethyl Corporation, 33, 1113
- Ethylene glycol, 117, 571
- Etiological investigation  
in epidemiology, 209–211  
prospective study application to, 245
- Etiological mechanisms, 284
- Etiological studies, 287
- Etiology, 1403, 1404–1405
- EU-OSHA (European Agency for Health and Safety at Work), 1579
- European Union  
chemical regulation by, 773  
children’s health issues, involvement in, 806
- environmental health involvement of, 1578–1579
- laboratory practice guidelines developed by, 191
- occupational and environmental standards of, 1561
- Eustress, 812
- Evaluation research, 258–259
- Evaluations, workplace, 1186
- Even odds, 251
- Evidence  
coherence of, 214  
standards of, 1417/<sup>t</sup>
- Evidence-based medical dispute resolution, 284, 1438–1441
- Evidence-based medical-expert practice, 41
- Evidence-based medicine  
occupational medicine role of, 207  
overview of, 1440
- Exacerbation *versus* aggravation, 1176–1177
- Examinations, workplace, adjustment of, 1186
- Exclusive causation, 1430
- “Exclusive remedy” (defined), 1320
- Excretion of xenobiotics, 66, 96–99
- Executive physical, 1212
- Exercise-induced asthma, 1070–1071
- Exit evaluations, 277
- Experimental association, 215
- “Expert medical examiner” (defined), 1331
- Expert witnesses  
physicians and health professionals as, 1265, 1390, 1397–1398, 1440  
role of, 1390–1391, 1397–1402  
skills involved in serving as, 1432–1438  
weight of evidence covered by, 1415, 1416

- Exploring the Dangerous Trades* (Hamilton), 28
- Explosives, terrorist use of, 1386, 1387
- Exposure
- biomarkers of, 262
  - circumstances of, 1019–1020
  - constant levels, assumption of, testing, 314–315
  - control of, 1202
  - defined, 108
  - extent among cases, 250
  - opportunity for, 880
  - protection need, determining, 296
  - records of, 272–273, 276, 1284–1285
  - safe levels, determining, 65
  - source, identifying, 296
- See also* Acute exposure guideline levels (AEGLs); Occupational exposure
- Exposure assessment
- accuracy of, 205
  - applications of, 319–323
  - approaches to, 303*t* (*see also* Direct assessment of exposure; Indirect assessment of exposure)
  - definition and overview of, 380–381
  - of environmental or occupational hazards, 739–740
  - means of exposure and, 66
  - standards, setting, 295
- Exposure-based studies, 233
- Exposure–effect relationship, 297
- Exposure monitoring
- lack of, reasons for, 276
  - personal, 273
- Exposure–outcome association, 228
- Exposure–response relationships
- association in absence of, 214
  - definition and overview of, 381
  - error probability, tolerable in, 217
- extrapolations fo, 396*f*
- See also* Epidemiological exposure–response relationship; Toxicological exposure–response relationships
- External healthcare facilities. *See* Community-based healthcare facilities
- Extrapolation, 313–314
- Extrinsic allergic alveolitis (term), 1080
- See also* Hypersensitivity pneumonitis
- Exxon Valdez tanker collision, Alaska, 1989, 1450
- Eye disorders, 1015
- Eye irritation, 754
- Eye protection, 873–875
- Face shields, 874–875
- Facilities for employees
- access barriers, removing, 1186
- Factor, 224
- Fair Housing Act, 1135
- False finding, 230
- Falsifiability, 211
- Family and Medical Leave Act (FMLA), 1993, 1156–1158, 1163
- Family physicians
- occupational medicine practiced by, 1261, 1263
  - on-site medical practitioners as, 1253
- Family practice
- occupational health services compared to, 1291
- Family stress
- workplace support lacking in cases of, 820
- Farm incidents
- predictors of, 370, 372
- Farmer’s lung, 1081, 1081*f*

- Fat-soluble xenobiotics  
concentration of, 80–81
- Fatalities, work-related, 963–964
- Fatty liver  
toxins causing, 120
- Fault, apportionment of, 1342
- Fault tree analysis, 881
- Federal Aviation Administration,  
1471–1472
- Federal employees  
workers' compensation for, 1322
- Federal Motor Carrier Safety Admin-  
istration (FMCSA), 1153
- Federal Pesticide Control Act, 1972,  
1448
- Federal Rules of Evidence, 1390
- Fee-for-service healthcare  
physician attitudes toward, 30
- “Fellow servant” doctrine, 1324
- FEMA trailers  
formaldehyde in, 714
- Fertility  
workplace exposure effect on, 1105
- Fetal alcohol syndrome  
ethanol associated with, 178
- Fetotoxicity  
organogenesis and, 179–180
- Fibers, 127
- Fibrogenic dusts  
reactions to, 1039–1041
- Fibromyalgia, 1006
- Fibrosis, 120–121
- Fibrous mordenite, 578
- Fibrous silicates, 578–580
- Fibrous silicates, non-asbestos, 577t
- Financial counseling, 1232
- Financial industries, 1280
- Financial markets  
risk science applied to, 380
- Fine particulate air pollution  
mortality associated with, 752
- Fine particulate matter  
as category, 756  
respiratory tract, behavior in, 133
- Finkelstein's test, 996*f*
- Finland, occupational healthcare in,  
32–33
- Finnish Institute for Occupational  
Health, 32–33
- Firefighters  
chemical carcinogens, exposure to,  
151
- fitness for duty evaluations for,  
1153
- infertility problems in, 1105
- meta-analysis in studies on, 222
- mortality studies among, 288–292,  
288*f*
- mutual employment protection of  
aging, 237
- “First, do no harm” (concept), 467
- First-aid reports, 277–278
- First-order elimination rate constant,  
103
- First-order kinetics  
conditions resulting in, 102  
overview of, 103–104
- First responders, 1371–1373
- Fisher, R.A., 217
- Fishing  
as high-risk occupation, 960
- Fit (defined), 1141
- Fit subject to work modification  
defined, 1141, 1142  
as pre-placement evaluation  
outcome, 1145–1146
- Fitness  
as health promotion component,  
1221–1222, 1228, 1230–1231
- Fitness at time of hire  
healthy worker effect stemming  
from, 236

- Fitness for duty (FFD)  
defined, 1136
- Fitness for duty (FFD) evaluations  
back pain or injury history, identifying through, 985–986  
designing, 1137–1141, 1136t  
outcomes, reporting, 1141–1142  
overview of, 1135–1137  
types of, 1140 (*see also specific type of evaluation, e.g.: Pre-placement evaluations*)
- Fitness for return to work. *See* Return-to-work evaluations
- Fitness programs  
as health promotion component, 1223–1224
- “Fitness to work” (term), 1136
- Flame-related burns  
prevalence and occupational association of, 968
- Flight-or-fight response, 815
- Florida  
workers’ compensation in, 1329
- Flow rates (Q-dot), 99
- Focal emphysema  
causes of, 1034
- Focal fibrosis  
causes of, 1034
- Fog index, 423t
- Follow-up care  
for injured or ill worker, 946
- Follow-up period  
in prospective study, 248–249
- Food and Agricultural Organization (FAO)  
environmental health involvement of, 1577
- Food and Drug Act, 1st, 62
- Food and Drug Administration, 1473  
establishment of, 62  
food inspection procedures of, 794
- Food-borne illness and outbreaks  
causes and overview of, 791–792  
food pathogen types, 790t  
investigation of, 791t  
preventing and minimizing, 792–795
- Food preservation and protection  
chemical pros and cons, 793  
strategies for, 792t
- Food safety  
overview of, 790–795  
toxicology of, 70, 72
- Food service inspection  
food protection through, 795
- Footwear, protective  
overview of, 875–876
- Foreign body carcinogenesis  
overview of, 165–166
- Foreign operations, 1559–1564, 1571
- Forensic medicine  
history of, 59, 62
- Forestry  
injury rates in, 952
- Formaldehyde  
in building environment, 711, 714–715  
health effects of, 93
- Formic acid  
formation and effects of, 570
- Fossmann, Sven P. (1911–1990)  
as occupational healthcare system architect, 31–32
- “Frequentist” predictions, 1424, 1426
- “Frequentist” probabilities, 1417t, 1422
- “Frequentist” statistics, 1344–1345
- Friction  
as skin disorder cause, 1030
- Fluorocarbons  
bans on, 1449
- Frostbite  
overview of, 553

- Frozen shoulder syndrome (adhesive capsulitis)  
overview of, 997
- Frye rule, 1391, 1395–1396
- Fume fevers  
overview of, 1084–1085
- Fumigation point, 750
- Function, 1180
- Functional assessment  
as diagnosis component, 1019
- Functional capacity  
impairment as obverse of, 1171
- Functional capacity evaluation  
overview of, 1168–1170
- Functional impairment  
degree, assessing, 1338
- Functional loss, 1180
- Fungicides, 642
- Furans  
molecular structure of, 596*f*  
toxicity of, 597–600  
types of, 595*f*
- Furniture, emissions from, 715
- “Gaia hypothesis,” 794
- Galen of Pergamum, 58
- Garbage disposals  
as waste disposal method, 771
- Gas chromatography/mass spectrometry (GC/MS)  
as drug testing technique,  
1194–1195, 1196
- “Gas eye” (keratoconjunctivitis)  
causes of, 602
- Gases  
absorption of, 76  
exposure to, 125–126, 133–136  
penetration and deposition of, 126  
*See also* Toxic gases
- Gases, volatile, 98
- Gastric neoplasms  
substances linked to, 123
- Gastrointestinal absorption, 75–76
- Gastrointestinal tract  
cancers of, 123  
toxin effects on, 121–123
- Gene expression, 156
- General Adaptation syndrome,  
815–816, 814*f*
- General practitioners  
healthcare plans, attitude concerning, 30
- General presumption of risk, 1430,  
1431–1432
- General preventive medicine  
defined, 1217
- Genetic epidemiology (term), 205
- Genetic predisposition to cancer  
carcinogen exposure and, 163  
as occupational cancer factor, 164
- Genetic states  
as susceptibility state type, 114
- Genetics  
environmental hazard exposure and,  
206
- Genitourinary cancers  
as occupational cancers, 1096
- Genomic epidemiology  
biomarkers of, 286  
methodology and limitations of, 205
- Genomics  
current emphasis on, 206
- Genotoxicity  
DNA damage associated with, 158*f*
- Germany  
occupational healthcare in, 26–27  
workers’ compensation in, 1326,  
1327, 1352, 1353–1354
- Glare, minimizing, 514
- Glasses, protective, 873–875

- Global atmospheric change, 799–801  
 Global change, health implications of, 44  
 Global health (term), 1536  
 Global occupational and environmental health  
     border health and migrating populations, 1564–1570  
     definition and overview of, 1535–1537  
 Global (term), 1535  
 Global violence  
     ecosystem damage as, 45  
 Global warming, 800  
 Globalization  
     as environmental health issue, 39  
     overview and impact of, 1540–1541  
     rise of, 1535  
     and trade, 1551–1552  
 Glomerulonephritis  
     toxins as cause of, 117  
 Glomerulus  
     toxin impact on, 116  
 Gloves, protective  
     overview of, 868–872  
 Glucuronyl transferase  
     chemical structure of, 89/  
     function and action mechanisms of, 88–89  
     metabolism of, 88–89  
 Glutathione peroxidase  
     function and action mechanisms of, 91  
     metabolism of, 91  
 Glutathione transferase  
     function and action mechanisms of, 90–91  
     metabolism of, 90–91  
 Glycol ethers  
     utilization and toxicity of, 571–572
- Glycols  
     definition and overview of, 569–570  
     exposure, routes of, 571  
     hydrocarbons compared to, 671
- Glyphosate  
     use and health effects of, 664–665
- Goggles, 874–875
- Goldsmith, John R., 751
- Golfer's elbow. *See* Epicondylitides
- Goniometer, 1181
- Good Laboratory Practice (GLP), 191
- Greeks, ancient, occupational medicine of, 14
- Green Hospitals and Health Care movement, 49
- Greenhouse gases, 799–800
- Gross domestic product (GDP)  
     (defined), 1541
- Groundwater  
     contamination of, 762–763, 764  
     decontamination of, 767  
     formation of, 757  
     management of, 758–759  
     movement of, 758–759, 756/  
 Group practices  
     attitudes concerning, 30  
     occupational health services as part of, 1261  
     pros and cons of, 1302–1305  
     *See also* Multispecialty group practices
- Guidelines (defined), 440
- Guillain-Barré syndrome  
     organophosphate-induced delayed neuropathy (OPDIN) compared to, 654–655, 1119
- Guinea pig ileum, 109
- “Gulf War syndrome,” 660
- Guthrie Health, 25–26

- Gynecomastia  
as adverse reproductive outcome, 173  
in oral contraceptive manufacture workers, 175
- H1N1 influenza, 1122, 1381
- H5N1 influenza (avian flu), 1381–1382, 1383–1384
- Haagen-Smit, Arie J., 751
- Haber, Fritz, 135–136
- Haber's law for toxic gases, 135–136, 135*f*  
nitrogen oxides, 633  
substances not following, 601
- Habituation, 60
- Haddon, William, Jr., 368
- Haddon's Matrix, 369, 371*t*, 474
- Haddon's Ten Countermeasures, 368*t*
- Hadler, Norton, 990
- Half-life  
defined, 102
- Halogenated hydrocarbons  
health effects of, 676  
liver affected by, 120  
transcutaneous absorption of, 77
- Halomethanes, 678  
legacy and values espoused by, 13  
mentors and influences in, 1204  
toxicology interests and expertise of, 64  
works of, 18
- Hand injuries  
prevalence of, 951, 958  
types of, 995–996
- Hand protection, 867–872
- Hand washing  
infectious disease spread, preventing through, 1127  
tips for, 868
- Handicap (defined), 1172, 1175–1176
- Hardy, Harriet, 28
- Harm, apportionment of, 1342
- Harrah's Entertainment, 1259
- Harrell, W.A., 370, 372
- Hat makers, 1114
- Havner-Robinson, 1396
- "Hawk's Nest" water tunnel disaster, 1930–1931, 1046
- Hazard Analysis and Critical Control Point System (HACCP), 794
- Hazard control  
overview of, 845–847  
strategies for, 847–851
- Hazard exposure  
controlling, 846*t*  
monitoring, 1278  
occupational breakdown for, 902*t*
- Hazard Ranking System, 787, 788
- Hazardous chemicals. *See* Chemical hazards
- Hazardous occupations  
safety measures, evolution of, 286  
socioeconomic status, low linked to, 225
- Hazardous substances  
disposal of, 781–782  
occupational health protection for hazardous materials workers, 1444
- Hazardous waste disposal  
overview of, 782–785  
persistent toxic substances (PTSs), 789–790  
requirements for, 769  
in small quantities, 781–782  
special wastes, 785  
superfund, 785–789  
uncontrolled, health consequences of, 771
- Hazards  
communication systems and standards, 425–426  
defined 876

- as disease cause, 43–44  
identification and characterization of, 380, 388–389, 392, 878, 880–881  
injuries induced by/as incident element, 358  
perception of, 406  
reappearance of old, 1016  
*See also* Occupational hazards; Workplace hazards; *specific type of hazard*
- Hazards, physical  
concepts of, 505  
public health triad applied to, 473*f*
- HAZOP (hazard and operability analysis), 881
- Headache, 1237
- Health  
concept defined, 463  
concepts of, 481–482  
employer questions about, 1149–1150  
in occupational context, 1277–1278, 1276*t*  
Health, safety and environmental affairs (HSE) department, 1274
- Health and safety committee, 1276–1277
- Health and Safety Executive (United Kingdom), 876, 885
- Health and wellness programs, workplace, 51  
absence, reducing through, 1161  
economic aspects of, 1208  
physician support of, 1202
- Health Belief Model, 490–491
- Health Canada, 380*f*, 382
- Health conditions, pre-existing, 818
- Health education  
in health evaluations, 1215, 1216  
at health fairs, 1211–1212  
as health promotion component, 1221, 1223, 1224, 1226  
for immigrants, 1204, 1570  
post-World War II, 1205  
preventive medicine role of, 1218  
screening program opportunities for, 1210, 1211, 1212  
in service industries, 1280
- Health evaluations  
executive, 1214–1216  
medical records as source of, 277  
periodic, 1209–1213
- Health fairs, 1211–1212
- Health hazard appraisals, 277
- Health hazard evaluation  
overview of, 273  
pros and cons of, 323–324
- Health insurance  
health promotion programs compared to, 1228  
introduction, post-World War II of, 1205, 1228  
reform, 1328, 1330–1331
- Health maintenance organizations, 25, 1247
- Health problem risk factors  
reducing, 1223–1224
- Health problems  
certification of personal, 1155–1156  
identifying, 1167  
progression, stopping, 1232
- Health problems on job  
healthy worker effect stemming from departure due to, 236–237
- Health professionals, referral to outside, 1236
- Health promotion  
case management coverage compared to, 1237–1238  
community-based programs, 1225  
definition and overview of, 477–481, 1221–1225  
employer financial gain through, 1240

- Health promotion (*continued*)  
historic context, 1202–1209  
in service industries, 1280  
social importance of, 1228  
term usage, 1201  
wellness and productivity management, role in, 1239  
*worksite (see Worksite health promotion programs)*
- Health psychology, 1206
- Health-related behavior, changing  
on community scale, 1206–1207  
primary prevention and, 465–466  
promotion of, 1201  
stages of, 498*t*–499*t*
- Health-related behaviors  
health promotion programs targeting, 1226–1227  
outcome of, 260, 261
- Health-related expenses  
conditions linked to greatest, 1237
- Health-related risks  
case finding for, 892
- Health risk appraisal instruments  
development of, 1206
- Health risk assessment  
questionnaires, 1210
- Health status  
baseline, 280  
improvement, factors in, 485–486
- Health surveillance  
periodic, overview of, 891  
protocols, designing, 284
- Health survey  
sick building syndrome assessment by, 706–707
- Healthcare costs  
controlling, 1239
- Healthcare services, general  
occupational health services as historical basis for, 1247
- occupational health services  
compared to, 1291–1292
- occupational health services  
separate from, 1289–1290
- physician, patient and stakeholder relationships in, 1245*f*
- “Healthy People” (U.S. Surgeon General and Office of Disease Prevention and Health Promotion), 1221–1222
- “Healthy worker effect”  
as prospective (cohort) study bias, 235–238, 285  
reference population pitfalls compared to, 247
- Hearing conservation  
overview of, 1085–1089  
programs, components of, 543–544, 541*t*
- Hearing loss or impairment, 1088–1089  
*See also Noise-induced hearing loss*
- Hearing mechanism  
energy conversion in, 1086, 1085*t*
- Hearing protection, 862–865
- Heart disease  
death from, 482, 836
- Heat stress  
overview of, 558–560  
prevention of, 560–562
- Heat syncope, 562–564
- Heating, ventilation, and air conditioning (HVAC) system  
maintenance and cleaning of, 735  
purpose of, 712, 713  
*See also Ventilation*
- Heinrich, Herbert, 366
- Helsinki Commission (HELCOM)  
composition and purpose of, 789
- Hemangiosarcoma  
toxins causing, 121

- Hemoglobin synthesis  
lead effect on, 617, 616*f*
- Hepatic acinus, 120*f*
- Hepatic angiosarcoma  
cluster investigations of, 238
- Hepatitis  
chemical and alcohol compared, 121
- Hepatitis A  
travelers, risks to, 1572
- Hepatitis B  
healthcare workplace risks of, 1121  
travelers, risks to, 1572
- Hepatobiliary excretion  
thresholds for, 97
- Hepatocellular damage  
toxins causing, 120
- Hepatocytes  
xenobiotics acting on, 119
- Hepatotoxic agents  
excretion and metabolism, interference with, 97–98
- Heptachlor  
use and toxicity of, 660
- Herbicides  
overview and toxicity of, 664–665  
use of, 642
- High altitude  
effects of, 549–550
- High-altitude sickness, 1571
- High-pressure neuroleptic syndrome, 546, 547
- High-priority hazards  
management of, 882, 884
- High-risk groups  
intervention for, 476, 477
- High-risk industries  
company size, typical in, 953  
overview of, 952–953
- High-risk occupations, 960–962
- High-technology industries  
worker injuries in, 1279–1280
- Hilar lymphadenopathy  
appearance of, 1042*f*
- Hill, Sir Austin Bradford, 212, 215–216
- HIPAA, 276, 329, 1513
- Hippocrates, 12
- Hiring practices  
Americans with Disabilities Act  
coverage of, 1133–1134
- Historical perspective (cohort) study  
case-referent studies compared to, 250
- defined, 244
- overview of, 246
- HIV/AIDS  
as chronic disease, 204  
in developing countries, 1538, 1555, 1563  
emergence of, 1126  
healthcare workplace risks of, 1120–1121, 1122  
prevention of, 1124  
travelers, risks to, 1572
- HMOs, 31
- “Holistic,” 41
- Holmes, Oliver Wendell, 483
- Homeland security legislation and policies  
community right to know *versus*, 429
- Homes  
building-associated outbreaks in, 689  
hazardous substance exposure in, 781  
humidity level in, 697  
radon in, 698
- “Honeymoon effect” as burnout factor, 833
- Hormone-induced carcinogenesis, 165, 166

- Hospital-based occupational health services  
clinics, overview of, 1260, 1261, 1267  
pros and cons of, 1302–1305
- Host defenses against toxins, 136–143
- Hot wire asthma, 1077
- Housekeeping  
as occupational hazard control strategy, 851
- Hugo, Victor, 18
- Hull House, 27
- Human Development Index  
(defined), 1541
- Human factors school of thought  
injury theories influenced by, 365
- Human Health and the Environment* (Chivian), 45
- Human populations  
complexity of, 230
- Human reproduction  
obstacles and adverse outcomes in, 172–174
- Human resources department  
occupational health and safety unit within, 1271
- Human rights law, 1132–1133, 1576
- Human studies  
in inhalation toxicology, 144, 148
- Humidity extremes  
in building environment, 696, 697
- Hunter, Donald (1898–1978)  
work and writings of, 34
- Hurricane Katrina  
FEMA trailers set up following, 714  
lessons from, 1361–1365  
response, perception of, 404
- Hurricane Rita  
lessons from, 1361–1365
- Hydrazine  
liver affected by, 120
- Hydrocarbon vapors  
in confined spaces, 1107  
exposure effects of, 1111
- Hydrocarbons  
health effects of, 676  
use of, 671
- Hydrofluoric acid (HF)  
skin effects and treatment for, 1023–1024
- Hydrogen sulfide  
acute exposure guideline levels (AEGLs), 456*f*  
exposure effects of, 1107–1108, 1109, 1111  
overview of, 600–605  
toxicity of, 110
- Hydrolytic reactions  
in toxicology, 94
- Hygiene survey  
sick building syndrome assessment by, 707–709
- Hypersensitivity pneumonitis  
appearance of, 1080*f*, 1081*f*  
causes of, 695  
fume fevers, 1084–1085  
industrial bronchitis, 1084  
overview of, 1080–1084
- Hypertension  
population cross-studies, 476–477
- Hypertrophic skin changes  
solvents as cause of, 1025*f*
- Hypochondriasis  
somatization compared to, 932
- Hypothermia  
causes and risk factors for, 552, 554  
treatment of, 554–556
- Hypothesis  
study design strengths, relative for proving, 257*f*
- Hysteria  
term, 932

- Illiteracy in workplace combatting, 425–426
- Illness attitudes concerning, 1166 prevention of premature, 460 work-related, absence due to, 1163
- ILO Classification of the Radiographic Appearance of the Pneumoconioses, 1038
- Imaginary Time Bomb, The* (Mullan), 1208–1209
- Immigrants health education for, 1204, 1570
- Immigration and Customs Enforcement (ICE) raids, 1570
- Immune and immune-related mechanisms in lung, 139–140
- Immunization ethical considerations, 467–468
- Immunization programs employee participation in, 1127–1128
- Immunizations, 1382–1384
- Impairment capability evaluation of persons with, 1170 defined, 1171, 1172, 1173–1174, 1175 temporary *versus* permanent, 1176
- Impairment assessment definitions and concepts, 1172–1177 outside workers' compensation, 1177–1179 overview of, 1170–1172 in workers' compensation, 1179–1183
- Impairment model of disability evaluation overview of, 1338
- Impulse meters (defined), 541–542
- In-plant health services hospital-based occupational health clinics influenced by, 1269 overview of, 1256–1257 scope of, 1252 types of, 1257, 1258–1259, 1260
- In vitro “Ames” assay as alternative to animal studies, 189 limitations of, 170, 193 overview of, 167
- In vitro studies in inhalation toxicology, 144, 148 in toxicology, 188, 189, 191, 192–194
- Incidence rates perspective studies yielding, 247
- Incident causes, theories concerning, 361–369, 362*t* definition and term usage, 357–358 response, 1357
- Incident command system (ICS), 1377–1378, 1376*f*
- Incineration as waste disposal method, 770
- Inclinometer, 1181
- Indemnification definition and overview of, 1332
- Independent medical evaluation (IME) guidelines of, 1150 overview of, 892
- Independent medical examination (IME) defined, 1333–1334
- “Independent medical examiner” defined, 1331 duties of, 1333–1334
- Indirect assessment of exposure, 310–319

- Individualized health care  
trend toward, 1207–1208
- Individualized risk factors  
health promotion programs targeting, 1227  
intervention, 1237  
medical records information concerning, 276, 278–279
- Indoor air pollution  
children, effect on, 802–803  
exposure, limiting, 755  
factors in, 747  
pollutants and sources, 713–716, 712<sup>t</sup>
- Indoor air quality  
Environmental Protection Agency involvement in, 1444, 1446  
guidelines for, 1444  
overview of, 712–720  
as sick building syndrome (SBS) factor, 721
- Inductively coupled plasma mass spectrometry (ICP-MS)  
overview of, 683  
utilization of, 687
- Industrial accidents  
as occupational stress type, 842
- Industrial applications, products with testing of, 188
- Industrial bronchitis  
overview of, 1084
- Industrial Health Fatigue Board, 365, 366
- Industrial hygiene  
evaluation, 728, 730  
survey example, 727<sup>t</sup>
- Industrial hygienists  
duties of, 1251, 1254  
hazard control role of, 845  
job description and duties of, 846–847  
staffing requirements, 1283
- Industrial Medicine and Surgery* (Mock), 28–29
- Industrial medicine clinics, 1260, 1268–1269, 1281  
*See also* Occupational health centers or clinics
- Industrial medicine (term), 12, 29
- Industrial Revolution  
chemical contamination during, 62  
child labor during, 1554  
emergency management since, 1358  
lead poisoning during, 605  
occupational medicine during, 19–24  
worker conditions during, 1203, 1324–1325  
worker safety, attitudes during, 363–364
- Industrial Survey Unit of Arthritis and Rheumatism Council, 989
- Industry sector  
lost-time claim rates by, 959<sup>t</sup>
- Industry-wide studies  
design and overview of, 233
- Infants  
nitrate effects on, 764
- Infectious disease  
chemical carcinogenesis compared to, 149  
community-acquired, prevention measures for, 1126–1128  
epidemiology and, 202  
public health triad application to, 471–473, 472<sup>f</sup>
- Infectious hazards in workplace  
overview of, 1121–1125
- Infertility  
prevalence of, 174
- Influenza  
travelers, risks to, 1572

- Influenza outbreaks  
  workplace impact and prevention of, 1127–1128
- Influenza pandemics  
  threats and preparedness, 1381–1382
- Information service industries  
  occupational health needs in, 1280
- Infrared  
  overview of, 536–538
- Ingestion  
  absorption compared to, 72  
  as environmental exposure route, 70  
  in occupational medicine, 70, 72
- Inhalation  
  absorption in, 75–76  
  absorption resulting from, 74  
  absorption secondary to, 76  
  ingestion resulting from, 72  
  as workplace exposure route, 70  
  xenobiotic absorption through, 76
- Inhalation toxicology  
  aerosol–gas mixtures, 136  
  aerosols, 126–133  
  experimental, 143–148, 145*f*  
  gases, 133–136  
  host defenses, 136–143  
  overview of, 123–126
- Initial distribution  
  pharmacology and toxicology application to, 83
- Initiation (cancer phase)  
  overview of, 155  
  requirements for, 158  
  threshold of, 156
- Injured workers  
  assessment of, 51  
  employer outreach, attitudes concerning, 1301  
  expectations of, 944  
  follow-up period, 953–958  
  historic conditions for, 1325
- history of, 954–958  
identity change of, 954  
legal rights of (*see* Workers' compensation)  
profile of, 1411  
support groups for, 958  
treatment of, 1250, 1265–1266, 1268, 1279, 1286, 1331–1332
- Injury  
  cumulative as legal doctrine, 1346–1347  
  disease *versus*, 1345  
  evaluation of, 944  
  incident investigation following, 965–967  
  medical *versus* workers' compensation definitions of, 949  
  prevention of, 1016  
  psychological dynamics of, 924–929  
  public health triad applied to, 474  
  pyramid, 358–361, 360*f*  
  susceptibility to, 358  
  *See also* Occupational injuries
- Injury claims  
  FFD evaluation reviewed in, 1137
- Injury prevention  
  overview of, 357  
  principles of, 359*t*, 370  
  scope of, 358  
  theories concerning, 366
- Injury reports  
  medical records including, 277
- Injury risk reduction  
  fitness program role in, 1230
- Insecticides  
  overview of, 642
- Insoluble gases  
  health effects of, 134
- Insurance  
  risk science applied to, 380

- Insurance and benefits  
cost shifting between, 1340–1341, 1352
- Insurance companies  
environmental medicine, policies concerning, 40
- Intermediate neurotoxicity syndrome (IMS)  
overview of, 655
- Internal medicine  
occupational medicine as subspecialty within, 9
- International Agency for Research on Cancer, 1406
- carcinogen classification system of, 151–152, 1095
- formaldehyde carcinogen status determined by, 714
- radiation-induced cancers identified by, 526
- shift work health effects studied by, 825–826
- International Association for Ecology and Health, 45
- International Atomic Energy Agency (IAEA)  
environmental health involvement of, 1577
- International Commission on Occupational Health (ICOH)  
code of ethics, 1514–1529  
overview of, 9, 1579–1580
- International Joint Commission (IJC), 1477  
toxic waste management role of, 789
- International Labour Organization (ILO), 885, 1038  
environmental health involvement of, 1577, 1578
- International Network for Children's Health, Environment, and Safety (INCHES)  
purpose of, 806
- International Occupational Hygiene Association, 876, 1580
- International Occupational Hygiene Organization, 885
- International occupational medicine/healthcare post-World War II, 31–33  
pre-World War I, 26–27
- International Organization for Standardization(ISO), 1561  
function of, 742–743
- International Programme on Chemical Substances (IPCS), 886
- International Review Board  
function of, 276
- International Social Security Association, 1579–1580
- International Society of Doctors for the Environment (ISDE), 48
- International Society of Infectious Diseases  
travel medicine information provided by, 1571
- International SOS, 1575
- International (term), 1535–1536
- Internists  
occupational medicine practiced by, 1261, 1263
- Interstitial fibrosis. *See Asbestosis*
- Intervention studies  
in occupational epidemiology, 234–235
- Intestinal excretion  
overview of, 98
- Intracellular enzyme systems  
substrates for, 83

- Intratracheal instillation in inhalation toxicology experiments, 147–148
- Intravenous injection toxicokinetics in, 100
- Inversion definition and overview of, 743–744
- Inversion layer definition and overview of, 749
- Ionizing radiation health effects of, 524–531 measurement and assessment of, 521–524 overview of, 518–519 regulation of, 529–531 sources and exposures, 519–521
- Iron quantity present in body, 680–681 toxic, gastrointestinal tract affected by, 122
- Irrigation waters disease exposure risks of, 763
- Irritant contact dermatitis, 1024
- Irritant-induced occupational asthma, 1077–1080
- Jet lag, 1575–1576
- Job Accommodation Network (JAN), 1187–1188
- Job change-initiated evaluations, 1142, 1152–1153
- Job descriptions documentation for, 282 fitness for duty evaluations, application to, 1139–1140, 1141t–1142t satisfactory, examples of, 1141t–1142t
- Job exposure matrix overview of, 315
- Job performance decline in, investigation of, 1148–1149
- following illness or injury, 1146
- trail basis assessment, fitness for duty evaluation based on, 1141
- Job-related stress. *See* Stress in workplace
- Job requirements functional capacity evaluation and, 1168
- Johnson Controls* case, 1105
- Journal of Occupational and Environmental Medicine*, 10
- Jungle, The* (Sinclair) writings of, 62
- Junk science, 1413
- Kaiser health plan origin/model of, 25
- Karasek, Robert, 813
- Karoshi phenomenon as burnout factor, 835–836
- Kehoe, Robert (1893–1992), 33
- Kelsey, Frances, 179–180
- Kennaway, Sir Ernest carcinogen studies of, 154
- Kepone as carcinogen, 660 historic use of, 661 intervention for, 97
- Kettering Laboratory, 33
- Kidney carcinogenicity in, 118 excretion role of, 96–97 storage function of, 82 toxin effects on, 97, 115–118 as xenobiotic target organ, 80
- Kidney stones melamine as cause of, 117
- Kinetic energy definition and components of, 505

- Kinetics  
of xenobiotics, 99–105
- Labor-intensive industries  
worker injuries in, 1280
- Labor-management relations  
occupational health services and, 1275–1277
- Labor relations department  
occupational health and safety unit within, 1271–1272
- Labor unions  
medical services and insurance provided by, 1301
- Laissez-faire attitude  
safety approach characterized by, 363
- Lakoff, George, 1501, 1506
- Lalond Report  
overview of, 487
- Landfills. *See* Sanitary landfills
- Large-scale environmental issues  
health issues, 3
- Larynx  
chemical carcinogens affecting, 152
- Last, John, 1217
- Latency period  
association identification challenge posed by long, 284–285  
in cancer, 157 (2XP: cancer latency), 181, 284–285
- Latex gloves, 868–869
- “Law of vicarious responsibility,” 1324
- Lead  
air concentrations of, 757  
bioavailability of, 69–70  
compounds, 606–608  
in drinking water, 765, 766  
historic and current use of, 605–606  
intake *versus* bodily fluid concentration of, 682
- kidneys affected by, 117  
neurological effects of, 1117–1118  
as pollutant category, 756, 772–773  
reproductive problems caused by, 174
- Rome, ancient, exposure in, 58–59
- Lead and Copper Rule (LCR)  
overview of, 766
- Lead toxicity and health effects  
in adults, 609, 612–615, 1013–1014, 1113  
anemia, 616–617, 619–620  
in children, 81, 609, 610–612, 754, 801, 804, 1113  
encephalopathy, 1113  
on hemoglobin synthesis, 616*f*  
historic perceptions, 33, 58–59, 63  
overview of, 609  
reactions, individual, differing in, 111  
treatment for, 615–616
- Leaded gasoline  
hazards and bans on, 754, 1113
- Leather gloves, 872
- Legal standard  
in risk management, 432
- Legge, Sir Thomas Morrison (1863–1932)  
aphorisms of, 24, 25  
Hamilton, Alice compared to, 27  
life and work of, 24
- Legionnaires’ disease  
as building-related disease, 694–695
- Legislated entitlement programs, 1428, 1429
- Leibniz, Gottfried, 18
- Lethal concentration  
defined, 186
- Lethal dose  
calculating, 190  
defined, 186

- maximum tolerated dose (MTD)  
compared to, 192
- Leukemia  
as occupational cancer, 1096, 1097
- Liability  
injured worker return to work,  
1147
- “Libby mineral.” *See* Asbestos
- Life expectancy  
concerns over, 481  
trends in, 482, 483–484
- Lifestyle  
coercive measures concerning,  
1203–1204  
education regarding healthy, 1223  
executive programs focusing on,  
1216
- Lifestyle changes  
health risk, individual, impact on,  
1201
- Lifestyle factors  
medical record information  
concerning, 278–279
- Lifestyle risk factors  
for cancer, 1090  
studies, preliminary of, 241
- “Lifetime health monitoring  
program” (LHMP)  
definition and overview of, 1212
- Lifting  
low-back pain associated with, 972,  
979–980  
safe practices in, 984–985
- Light duty  
defined, 955–956
- “Light duty” (term), 1142, 1147
- Light in visible spectrum  
overview of, 534–536
- Line management  
occupational health and safety unit  
within, 1272–1273
- Linguistically isolated workers  
health promotion program  
influence limited for, 1227
- Linkage (between sets of data)  
overview of, 216  
ruling out, 228
- Lipid-soluble substances  
absorption of, 73–74 (lipophilic or  
hydrophobic)  
excretion of, 99
- Liquid wastes  
management of, 783
- Litigation  
certainty role in, 1412–1423  
social functions of, 1391–1393  
tort, 1393–1397
- Liver  
as metabolic site, 82  
toxin effects on, 118–121  
toxins, action on, 97–98  
as xenobiotic target organ, 80
- Living organisms  
uptake by, 797
- Local Emergency Planning Committee (LECP), 1380
- Location assignments  
documentation for, 282
- Locus of control (concept)  
overview of, 491–492
- LOEL (lowest observable effect level)  
definition and overview of, 186–187
- Logistics  
in incident command system (ICS),  
1377–1378
- London “killer fog,” 742
- Long-term disability (LTD)  
overview of, 1178  
*versus* workers’ compensation, 1179,  
1337, 1340–1341, 1352
- Longshoremen and Harbor Workers’  
Act, 1927, 1322

- Los Angeles  
air pollution in, 744, 751
- Loss of earnings capacity in disability evaluation  
overview of, 1338
- Low-back injury  
mechanics of, 979–980
- Low-back pain  
ability to work despite, 971  
clinical indicators of nonorganic, 983*t*–984*t*  
economic factors in, 971–972  
factors associated with, 951  
musculoskeletal disorders, other compared to, 970, 971  
overview of, 972–987  
pathology in, 973*t*–974*t*  
prevention of, 981*t*–982*t*, 987–988  
workers at risk for, identifying, 985–986  
workers' compensation claims for, 1335–1336  
*See also* Acute low-back pain;  
Chronic low-back pain
- Low-back x-rays  
in fitness for duty evaluations, 1139
- Low-priority hazards  
management of, 884
- Low-risk industries  
overview of, 952
- Lower class  
impoverishment of, 22–23
- Lower respiratory tract  
infections of, severity, factors affecting, 141–142  
particle deposition in, 131, 1032*f*  
upper respiratory tract infection impact on, 143
- Lower-technology industries  
worker injuries in, 1279–1280
- Lowest observed adverse effect level (LOAEL)  
defined, 443
- Lucent Technologies  
cardiovascular risk factor screening by, 1237
- Lung  
anatomic structure of, 124–125, 124*f*, 130  
chemical carcinogens affecting, 152–153  
defense mechanisms of, 125, 136, 1032  
excretion role of (*see* Respiratory excretion)  
injury, expressions of, 1032–1035  
particle deposition in, 76  
as xenobiotic target organ, 80
- Lung cancer  
air pollution as factor in, 754  
asbestos-related, 574–575, 584, 1098–1099, 1098*f*  
causes of, 1403–1404  
genetic predisposition to, 164  
as occupational cancer, 1095, 1096–1097  
occupational causes of, 35, 154, 211  
rate of, 263  
risk, relative of, 221  
risk factors for, 231, 1050  
screening and detection of, 162  
silicosis and, 1048*f*  
smoking linked to, 154, 211–212, 227, 231, 584, 1098*f*  
socioeconomic class, low, link to, 227
- Lung disease  
pathogenesis of, 133
- Lung diseases, occupational  
cause, establishing in, 1407–1408  
diagnosis of, 1037–1038

- evaluation of, 1035–1036  
in Middle Ages, 14–15  
studies on, 35
- Lung disorders  
prevalence of, 1015
- Macrophages  
respiratory effects of, 1033–1034
- Magnesium  
brain as target of, 83  
quantity present in body, 680–681
- Magnetic fields  
electrical fields, relationship to, 513f  
electromagnetic spectrum and, 514–518
- Maimonides, 60
- Mainstream cigarette smoke, 716
- Maintenance, heating and air conditioning mechanic, 1141t–1142t
- Maintenance (as behavioral change stage), 496
- “Major life activity” (defined), 1134
- Malaria  
occupational exposure and prevention, 1125–1126  
traveler risk and prevention, 1572, 1573–1574
- Malingering  
overview of, 935–937
- Malpighi, Marcello, 18
- “Man down” (term), 1109
- Managed care  
in occupational medicine, 1310–1313  
roots of, 1205  
workers’ compensation, 1350
- Managed care organizations (MCOs)  
marketing by, 1311  
primary care provided by, 1310  
rise of, 31
- Mandatory referrals  
to employee assistance programs, 1234
- Manganese  
intake *versus* bodily fluid concentration of, 682  
neurological effects of, 1112, 1115, 1114f
- Manufacturing  
injury rates in, 952
- Maquiladora* system, 1567–1568
- Maritime workers  
workers’ compensation for, 1322
- Markers  
clinical conditions serving as, 317t  
definition and overview of, 317–318
- Marketing cube, 1292, 1294–1300, 1293f, 1295f, 1296f, 1297f, 1298f
- Marketing in occupational medicine  
overview of, 1286–1287  
principles of, 1287–1290  
representation and communication, 1290–1292  
strategy, 1292, 1294–1301
- Marmot, Michael  
British civil servants studied by, 814  
population health model, role in developing, 487, 488
- Marxist-Leninist socialism, collapse of, 1542
- Mass hysteria (term)  
psychogenic outbreaks described as, 721
- Massachusetts  
workers’ compensation in, 1329
- Material data safety sheets (MSDSs)  
defined, 425
- Maternal toxicity  
fetal effects of, 177–178
- “Maximum medical improvement” (MMI) (term), 1176

- Maximum tolerated dose (MTD)  
determining in carcinogenesis  
bioassays, 167, 191–192
- McCord, Carey  
toxicology interests and expertise  
of, 64
- McCready, Benjamin  
public health reform efforts of, 23  
writings of, 18, 22
- Measurement and sampling  
theory of, 299–302, 301*t*
- Meat wrapper's asthma, 1077
- Media (air, water, and land)  
as ambient environment division, 4
- Media communication  
practical skills for, 419
- Media pollution (defined), 772
- Medical care, general. *See* Healthcare systems, general
- Medical certification  
absence, requirement after, 1166  
of personal health problems,  
1155–1156
- Medical Disability Advisor*, 945
- Medical dispute resolution. *See*  
Evidence-based medical dispute  
resolution
- Medical evaluations  
Americans with Disabilities Act  
coverage of, 1133, 1134  
functional capacity evaluations  
compared to, 1169  
guidelines for, 1150, 1151–1152  
historic practices, 1136  
outside employee assistance  
programs, 1236  
results, reporting, 1142  
for safety-sensitive positions,  
1154–1155  
*See also* Independent medical evaluation (IME)
- Medical fee schedules  
in workers' compensation, 1351
- Medical geographers  
studies conducted by, 241
- Medical group practices  
advocacy of, 30
- Medical history  
occupational and environmental  
history as component of, 893
- Medical inquiries  
grounds for, 1148–1149  
overview of, 1142  
pros and cons of, 1152
- Medical knowledge  
applications, non-medical of,  
1438–1439
- Medical leave  
law and legislation covering,  
1156–1158
- Medical model school of thought (in  
environmental medicine)  
overview of, 42–43
- Medical records  
access to, 1284–1285  
content and maintenance of,  
1250–1251, 1284–1286  
disease cause, establishing, role in,  
1408  
in epidemiological research, 272,  
276–281  
exposure history information  
limited in, 1407
- Medical review officers (MROs)  
drug testing role of, 1190–1193,  
1196  
job requirements and standards for,  
1191
- Medicine as social science, 1484–1485
- Medicolegal services  
medical expert role in, 41  
overview of, 1389–1391

- Melamine  
kidneys affected by, 117–118
- Men  
reproductive problems in, 173–174, 176
- Meningococcal disease  
travelers, risks to, 1572–1573
- Mental disorders  
interventions, employer role in, 1239  
productivity costs and expenses of, 1237
- Mental health  
components of, 819
- Mental illness and disorders  
employee assistance programs  
addressing, 1233
- Mercury  
health effects of, 1114  
kidneys affected by, 117  
overview and toxicity of, 620–624  
reproductive problems caused by, 174
- Mercury poisoning  
clinical descriptions of, 63
- Meredith, William, 1327
- Mesothelioma  
appearance of, 1100f  
asbestos linked to, 574, 575, 1098, 1101  
as occupational cancer, 1096  
overview of, 1099–1103
- Meta-analysis  
description and overview of, 222–223
- Metabolism  
biotransformation, environmental  
compared to, 798
- Metabolism of xenobiotics  
errors in, 163–164  
mixed function oxidase system, 85–88  
overview of, 66, 83–85  
reducing system, 94–96  
*See also* *Term metabolism of under specific substance, e.g.: Glucuronyl transferase: metabolism of*
- Metal fume fever, 1084, 1085
- Metal toxicity  
testing and diagnosing, 110–111
- Metallic mercury, 1114
- Metalloids. *See* Trace elements
- Metals  
as carcinogens, 152  
neurological effects of, 1117–1118  
physiological effects of, 81  
pneumoconioses caused by, 1037  
reproductive problems caused by, 174  
*See also* Trace elements
- Metals, toxic  
gastrointestinal tract affected by, 122  
kidneys affected by, 116–117
- Metastasis (as cancer phase)  
overview of, 156  
preconditions for, 163
- Methanol  
overview and effects of, 570–571
- Methyl-*n*-butyl ketone, 1118
- Methyl-parathion pesticide  
toxicity of, 649
- Methyl-tertiary-butyl ether (MTBE)  
water contamination with, 764
- Methylene dianiline  
liver affected by, 121
- Meuse Valley (Belgium)  
air pollution in, 744
- Mexican immigrants to United States, 1569–1570
- Mexico  
economic conditions in, 1566  
health issues in, 1566–1567, 1568–1569

- Mexico (*continued*)  
occupational medicine and health protection in, 1247, 1568
- Mice  
crowded populations and, 814  
life span of, 189
- Microbial hazards  
in water, 761–763
- Microsomal ethanol oxidizing systems (MEOS)  
metabolization of, 93–94
- Microwave  
overview of, 538–539  
radiofrequencies below, 539–540
- Middle Ages  
occupational medicine during, 14–16
- Middle-income countries (defined), 1545
- Middle-priority hazards  
management of, 882, 884
- Miettinen, Olli  
methodological research by, 256
- Migraine  
productivity costs and expenses of, 1237, 1239
- Migrant Clinicians Network (MCN), 1570
- Migrating populations, health issues affecting, 1564–1570
- Military physical examinations  
functional capacity evaluation, 1170  
occupational screening tests  
modeled after, 328  
as screening tool, 1145
- Mine Enforcement and Safety Administration (MESA), 1467
- Mine rescue personnel  
medical evaluations, regular for, 1155
- Mine Safety and Health Administration (MSHA), 1464–1468
- Miners  
lung cancer in, 154
- Miners' phthisis  
clinical descriptions of, 63
- Mines  
carcinogenic chemicals in, 153
- Mining  
child labor in, 1500  
health and safety hazards in, 14–15, 1465–1466  
historic conditions, 1499–1501  
occupational and preventive medicine in, 14
- Minority groups  
leadership roles as stress source for, 820–821
- Misclassification error, 223–224
- Les Misérables* (Hugo), 18
- Mixed function oxidase system  
metabolite activation in, 84  
overview of, 85–88  
oxidizing reactions of, 86f
- Mock, Harry E. (1880–1959), 28–29, 1204–1205
- Models, exposure level prediction through, 315
- Modena, 17, 18
- Modern Epidemiology* (Rothman), 231–232
- “Modified duty” (term), 1147, 1185
- Modifying factor (MF) (defined), 443
- Mold  
as allergen, 700  
conditions promoting, 697  
mycotoxins and, 624–627
- Mold illness  
questionable diagnosis of, 938
- Molecular epidemiology  
environmental medicine impacted by, 1481–1482
- Molecular epidemiology (term), 205

- Molecular oncology  
epigenetic carcinogenesis and,  
164–165
- Monitoring  
defined, 327
- Monitoring and surveillance  
concepts compared, 891  
disorder, course of, 333f  
ethics and social issues, 345–347  
for individual risk, 329–334  
objectives of, 330–332  
principles of, 332–335  
programs, implementing, 340–342,  
344–345
- Monodispersed aerosols, 128
- Montreal Protocol on Substances That  
Deplete the Ozone Layer,  
800–801
- Moral Politics* (Lakoff), 1501
- Morbidity  
determining, 260–261  
rates, decline in, 1145
- Morbidity ratio  
calculating, 268
- Morgagni, Giovanni, 18
- Mortality  
age breakdown for, 484–485  
concepts of, 481  
determining, 260  
reduction in, 482–485, 486
- Mortality, prospective  
studies on, 287–292
- Mortality ratio  
calculating, 268  
*See also* Proportionate mortality  
ratio (PMR); Standardized mor-  
tality ratio (SMR)
- Mosquito control and repellants,  
1125–1126
- Motor neuropathies  
overview of, 1117
- Mouthpiece studies  
in inhalation toxicology, 144, 148
- Movement disorders  
in animals, 183  
toxic exposures causing, 1115, 1114f
- Mt. Sinai Hospital  
occupational medicine program at, 35
- Mucociliary transport  
respiratory tract protection role of,  
137, 139  
viral infection impact on, 143
- Mucosal irritation  
air pollution as factor in, 754
- Mucous membranes, absorption  
through, 77
- Mullan, P., 1208–1209
- Multicompartment (higher-order)  
toxicokinetics  
overview of, 104–105
- Multicompartment kinetics model  
overview of, 102–103
- Multifactoral/multifactorial (defined),  
229
- Multilateral agencies, 1476–1478
- Multiphasic health screening  
definition and overview of,  
1210–1211
- Multiphasic screening  
emergence of, 1205
- Multiple chemical sensitivity (MCS)  
attitudes and beliefs concerning,  
937–941  
belief in, 46  
term, 42, 938
- Multiple-employer community-based  
healthcare services  
overview of, 1257, 1260–1261  
types of, 1261–1269
- Multiple-employer internal health  
services  
overview of, 1257, 1259–1260

- Multiple-employer occupational health services  
advantages of, 1289  
areas of emphasis of, 1300
- Multiple impairments  
rating of, 1182
- Multispecialty group practices  
overview of, 1260, 1266–1267
- Municipal solid waste. *See* Solid and municipal waste
- Musculoskeletal diseases  
term, 949
- Musculoskeletal disorders  
chronic, 990–991  
injury compared to, 949  
overview of, 950–951, 970–972  
*See also* Repetitive strain injury
- Musculoskeletal system impairment  
categories  
assessment within, 1181–1182
- Mutagenicity  
*in vitro* testing for, 192–194
- Mutual employment protection  
of aging workers, 237
- Mycotoxins and mold  
health effects of, 624–627
- Myocardial infarction  
beta-blocker effect following, 222  
high-demand job as predictor of, 813  
risk, relative of, 222
- Myofascial pain  
overview of, 1006
- N*-hexane, 1118
- N*-nitrosodimethylamine (NDMA)  
water contamination with, 767
- N95 and other limited protection devices  
overview of, 857–858
- Nanomaterials  
health effects of, 627–630  
water contamination with, 768
- Nanotechnology  
definition and overview of, 627–628
- Nasal cavity  
chemical carcinogens affecting, 153
- Nasopharynx  
lung protection role of, 136–137
- Nation (defined), 1542–1543
- National Academy of Sciences, 970
- National Ambient Air Quality Standards (NAAQS)  
overview of, 756–757, 754*t*
- National Back Fitness Test, 986
- National Commission on State Workmens' Compensation, 1329
- National Environmental Monitoring Inventory  
purpose of, 773–774
- National Environmental Protection Act (NEPA), 1970, 1443, 1447, 1448, 1453
- National health insurance  
attitudes concerning, 30, 31  
debate over, 1328
- National health insurance campaign  
in mid-twentieth century, 29–31
- National Health Service (United Kingdom), 27, 1326, 1328
- National Institute for Occupational Safety and Health (NIOSH)  
9/11 response role of, 1360  
environmental medicine theories espoused by, 42–43  
formation and role of, 11, 34–35, 1454  
productivity concerns addressed by, 1209  
records requested under authority of, 1285  
research functions of, 1465, 1466

- National Institute for Working Life (Sweden), 32
- National Institute of Drug Abuse (NIDA), 1190, 1191, 1192, 1193, 1196
- National Institute of Environmental Health Sciences children's health issues, involvement in, 806
- National Institute of Environmental Health Sciences (NIEHS) overview of, 11
- National Institute of Occupational Health (Sweden), 32
- National Institute on Alcoholism and Alcohol Abuse, 1232
- National Mortality Data Base, 290
- National Priority List (NPL), 787, 788
- National Research Council biomarker types recognized by, 196–197
- National Toxicology Program, 170, 182
- Natural aerosols overview of, 128
- Natural disasters lessons from past, 1361–1365 perception of and recovery from, 403–404 response to, 1357
- Natural environment as ambient environment division, 4
- Near-misses as occupational stress type, 842 reporting, value of, 369–370
- Neck injuries prevalence of, 951 syndromes, 992–995
- Negative stress (term), 813
- Negative studies evaluating, 220 publishing obstacles of, 223
- Neoplasm (term), 1433
- Neoplastic growth and proliferation in cancer, 156
- Nephrotoxic xenobiotics, 97, 115–118
- Nephrotoxicity animal experiments in, 118
- Neuropathic pain overview of, 1006
- Neurotoxins organophosphates and carbamates compared to, 646, 648
- Neutra, Ray cluster investigation work of, 239
- New Directions program, 1456
- New Zealand personal injury compensation system in, 1352
- NGOs (defined), 1543
- Nickel compounds, 682 as trace element, 681
- Night workers health and injury risks to, 828–829, 830
- Nitrate health effects of, 764 as pollutant category, 772 water contamination with, 764
- Nitric oxide overview of, 630
- Nitrile gloves, 872
- Nitriloacetic acid (NTA) carcinogenic properties of, 165, 166
- Nitrogen oxides of, 630–637, 755 toxic gases, comparison to, 141
- Nitrogen dioxide health effects of, 635–636 inhalation, pathophysiology of, 631*f* overview of, 630

- Nitrogen narcosis  
    overview of, 546–547
- Nitrogen oxide  
    overview of, 630
- Nitropropane  
    liver affected by, 120
- Nitrosamines  
    liver affected by, 120
- Nitrous oxide  
    overview of, 630
- No observed (adverse) effect level (NOAEL)  
    defined, 443
- Nobel, Alfred  
    blasting cap invented by, 632
- NOEL (no observable effect level)  
    definition and overview of, 186–187  
    maximum tolerated dose (MTD)  
        compared to, 192
- Noise  
    definition and overview of, 540–544  
    exposure, documentation for, 273  
    levels, examples of, 540t
- Noise dosimeters (defined), 542
- Noise-induced hearing loss  
    audiometric configurations for, 1087f  
    diagnosing, 1089  
    overview of, 1086–1088  
    prevention of, 543, 1085
- Noise reduction rating (NRR), 863–864
- Noncompliance  
    with safety standards, 446f
- Nondifferential misclassification  
    effect of, 224
- Nongovernmental organizations (NGOs), 1543, 1579
- Nonionizing radiation  
    infrared, 536–538  
    light in visible spectrum, 534–536
- microwave, 538–539  
    overview of, 531  
    radiofrequencies below microwave, 539–540  
    ultraviolet radiation, 532–534
- Normal states  
    impairment *versus*, 1174
- North American Agreement on Environmental Cooperation, 1478
- North American Free Trade Agreement (NAFTA), 1478, 1566
- Nose  
    avoidance behavior role of, 137  
    lung protection role of, 136  
    particle deposition in, 132  
    particle inhalation through, 130
- Nose-only exposure  
    in inhalation toxicology experiments, 144, 147
- “Not in my back yard” (NIMBY) syndrome  
    landfill site selection affected by, 768–769
- “Note from the doctor,” 1153, 1156
- Nuclear explosions, 1387, 1387
- Nuclear industry  
    workers’ compensation in, 1322–1323
- Nuclear Regulatory Commission  
    drug screening by, 1197  
    overview of, 1472–1473  
    radiation regulations of, 531
- Nuclear war  
    global violence and, 45
- Nuns  
    breast cancer in, 154
- Nutrition  
    education and counseling in, 1223, 1227  
    as health promotion program component, 1226, 1231

- working-class programs in, 1203
- Nutritional deficiencies  
epidemiology and, 202
- Obama, Barack, 1458
- Obesity  
social attitudes concerning, 1224
- Obsolescence  
in occupational epidemiology,  
286–287
- Occupation-related risk  
studies of, 233
- Occupational and environmental  
health research  
overview of, 893
- Occupational and environmental history  
applications of, 892*t*, 897–898, 900,  
905–906, 904*t*  
exposure inventory, 897*t*, 899*t*–902*t*  
form, 896*t*, 897  
levels of, 894–897, 892*t*  
overview of, 893–894  
structure of, 898*t*  
supplemental questions to,  
904*t*–906*t*
- Occupational and environmental  
medicine  
definition and scope of, 5, 12–13,  
1481–1482  
as field of practice, 7–11  
future of, 49–53  
history of, 3  
overview and integration of, 3–8,  
37, 38–39  
pyramid model of, 59
- Occupational asthma  
irritant-induced, 1077–1080  
out-migration of workers with,  
236–237  
overview of, 1068–1072
- pulmonary function, 1074*f*  
sensitizer-induced, 1072–1076,  
1071*t*
- Occupational cancer  
carcinogenesis and, 149  
epigenetics and, 165  
historic research on, 154  
overview of, 1091–1098  
research, history of, 154  
risk, factors influencing, 163, 164
- Occupational categories  
studies on, 234
- Occupational diseases and disorders  
burden, determining, 233  
case finding for, 892  
causation in, 1019  
clinical assessment in, 907*t*  
clinical descriptions, early of, 63  
clinical evaluation of, 908–923  
definition and overview of,  
1017–1018  
detection of, 64  
diagnosing, 1019  
distribution of, 1014–1015, 1012*f*  
evaluating, 1018–1019, 1409  
investigation, study designs for,  
221–222, 221*t*  
psychological aspects of, 923–937  
risk-modifying factors in, 908*t*–910*t*  
workers' compensation for,  
1345–1349
- Occupational epidemiology  
causation in, 209  
challenges of, 284*t*  
control groups in, 247  
environmental epidemiology,  
relationship to, 232  
exposure and response, measure and  
category of, 216  
limitations of, 283–287  
overview of, 206, 207–208

- Occupational epidemiology (*continued*)  
records used in, 274*t*–275*t*  
study design in, 232–238
- Occupational exposure  
assessment of, 298–299  
monitoring, 297, 324–325  
opportunity for, 320–322, 321*f*  
risk estimates associated with, 1344  
standards and guidelines, 442, 445
- Occupational hazards  
controlling, 847–851  
environmental hazards compared to, 739–740  
of first responders, 1371–1373  
*See also* Workplace hazards
- Occupational health  
environmental health overlapping with, 4–5  
global economics of, 1551–1552  
global organizations for, 1579–1580  
historic conditions, 1497–1501  
*versus* occupational safety, 1277  
problem solving, technical  
    approaches to, 844*t*  
and safety standards, 445, 447–552  
    (guidelines: 448–452)
- Occupational Health Advisory Committee  
establishment of, 1361
- Occupational Health and Disaster Expert Network (OHDEN), 1361, 1364
- Occupational health and injury tort litigation, 1394–1395
- Occupational health and safety  
in Canada, 1474–1476  
corporate medical department role in, 1270, 1271–1273  
foreign operations, 1559–1564  
future trends in, 1529–1535  
overview of, 1277–1279  
policy on, 1274–1275  
units, placement of, 1271–1273
- Occupational Health and Safety Act, 1970, 12, 34–35, 280, 445
- Occupational health and safety departments  
overview of, 1245
- Occupational health and safety practice priority setting  
as occupational epidemiology application, 284
- Occupational Health and Safety Review Commission, 1454
- Occupational health centers or clinics  
overview of, 1260, 1261, 1268–1269
- Occupational health interventions  
checklist of, 914*t*  
overview of, 915–923
- Occupational health monitoring  
overview of, 891
- Occupational health nurse practitioners  
duties of, 1253
- Occupational health nurses  
duties of, 1251, 1253  
importance of, 1282  
staffing requirements, 1283
- Occupational health professionals  
overview of, 1251–1255
- Occupational health professions  
scope of, 1250*t*
- Occupational health service (OHS)  
facilities  
design and supplies of, 1281–1282  
location of, 1279–1281, 1287–1288, 1289  
and operations, 1279–1305  
overview and types of, 1255–1257  
*See also under specific type of facility*,  
e.g.: Community-based healthcare facilities

- Occupational health services (OHSs)  
absence monitoring by, pitfalls of, 1159–1160  
core functions of, 1291*t*–1292*t*  
current trends in/revival of, 37–38  
definition and overview of, 1245–1246  
in developing countries, 1553–1554, 1562  
distribution of, 1247*t*  
economic viability of, 1286–1287  
emergency management role of, 1366–1376  
marketing of, 1286–1301  
physician, patient and stakeholder relationships in, 1247–1248, 1246*f*  
pyramid of, 50*f*  
quality *versus* price in, 1288  
record keeping (*see* Medical records)  
scope of, 1248–1250, 1247*t*  
staffing of, 1282–1284
- Occupational health trends  
surveillance for, 348–353
- Occupational healthcare system  
characteristics of, 1247–1251  
overview of, 1246–1247
- Occupational hearing conservations  
audiologists compared to, 1255
- Occupational hygiene  
measurement and sampling theory  
applied to, 301*t*  
organizational responsibility for, 1277  
techniques of, 303–308
- Occupational hygiene surveys  
data collecting for, 273  
overview of, 272
- Occupational hygienists. *See* Industrial hygienists
- Occupational infections  
community-acquired infection, 1126–1128  
infectious hazards in workplace, 1121–1125  
overview of, 1119–1121  
work-related infectious disease, 1125–1126
- Occupational injuries  
absence due to, 1163  
body part breakdown for, 951, 958–959, 959*t*  
clinical management of, 953–958  
complications of, 1333  
industries producing, 1280  
large-scale patterns of, 951–953  
overview of, 950, 951  
workplace scale of, 958–969
- “Occupational justice” model, 814
- Occupational medicine  
attitudes concerning, 1501–1507  
clinical approach to, 1016–1020  
as critical science, 1489–1497  
definition and scope of, 5, 6–7  
development and status of, 6  
environmental medicine compared to, 40, 53  
ethics in, 1507–1529  
origin and history of, 11–38  
post-World War II attitudes toward, 30–31
- Occupational Medicine Practice Guideline*  
series (American College of Occupational and Environmental Medicine). *See* ACOEM Practice Guidelines (APG)
- Occupational neurology  
overview of, 1112–1119
- Occupational physicians  
in Depression, 29

- Occupational physicians (*continued*)  
environmental health issues,  
    involvement with, 38  
post-World War II, 33  
rise in, 37  
training and qualifications of, 9–10
- Occupational safety  
    *versus* occupational health, 1277
- Occupational Safety and Health Administration (OSHA)  
9/11 response role of, 1360  
enforcement role of, 1459–1462  
exposure records keeping, standards  
    for, 272  
formaldehyde limits set by, 715  
formation of, 34, 1447  
function of, 445  
hazard communication standards  
    developed by, 425  
history of, 1452–1459, 1468  
Hurricane Katrina, response following, 1363  
immigrant outreach by, 1570  
legacy of, 1463–1464
- Mine Safety and Health  
    Administration (MSHA) separation from, 1464  
noise standards developed by, 543  
occupational exposure standards  
    developed by, 135  
occupational health hazards, authority of, 1473  
overview of, 11, 1452  
pesticide regulation by, 1451  
radiation exposure standards developed by, 529  
records keeping role of, 1513–1514  
records requested under authority of, 1285  
standards of, 447–448, 1472  
state programs, 1462–1463
- workers protected by agencies other than, 1471
- Occupational toxicology  
    ingestion in, 75–76  
    overview of, 63–65
- Odds ratio  
    in case-referent studies, 251, 252
- Odds ratios  
    overview of, 270–271
- Odor  
    building-related diseases, role in, 711–712
- OEM physicians  
    corporate medical department, role within, 1273  
    emergency management role of, 1368–1370, 1374–1375  
    environmental health involvement of, 740–742  
    focus, areas of, 7–8  
    liability position of, 1312–1313  
    prevention and risk assessment role of, 1013–1014  
    types of, 1251–1253
- Office buildings  
    building-associated outbreaks in, 689  
    temperature and humidity inside, 696–697, 698
- Office of Children's Environmental Health  
    purpose of, 804, 806
- Office of Disease Prevention and Health Promotion  
    school of thought of, 1221–1222
- Office operations  
    occupational health needs in, 1280
- Office workstations  
    ergonomics of, 512–514
- Offices  
    as low-risk environment, 4

- Oil and gas industry  
risk management in, 876t
- Oil companies  
medical services, historic role in, 1247
- Older persons  
air pollution exposure, limiting, 755
- Older workers  
safety, attitudes concerning, 372–373
- On Airs, Waters, and Places*  
(Hippocrates), 12
- On-site medical practitioners  
duties of, 1253
- Oncogenes  
expression activation of, 159
- One-compartment toxicokinetics  
model  
overview of, 103–104
- “One-hoss shay” (end of life  
concept), 483
- Operations  
in incident command system (ICS),  
1377–1378
- Opiates  
respiratory tract affected by, 142
- Options development  
definition and overview of, 381
- Oral cavity cancers  
substances linked to, 123
- Oral contraceptive manufacture workers  
gynecomastia in, 175
- Orf (spirochetal disease)  
as occupational hazard, 1124
- Organ system impairment categories  
assessment within, 1181
- Organic chemistry  
development of, 62
- Organic mercury, 1114
- Organization for Economic Cooperation and Development  
laboratory practice guidelines developed by, 191
- Organized medicine  
attitudes toward, 52
- Organochlorine insecticides  
molecular structure of, 655f  
toxicity and treatment for, 656,  
658–661
- Organogenesis  
developmental toxicity in, 179
- Organohalides  
child exposure to, 99  
elimination of, 104  
system-catalyzing reduction and  
dehalogenation in, 94
- Organophosphate esters  
neurological effects of, 1118–1119
- Organophosphate-induced delayed  
neuropathy (OPDIN)  
overview of, 654–655
- Organophosphate pesticides  
encephalopathic changes due to,  
1115  
molecular structure of, 646f  
neurological effects of, 1118–1119  
toxicity and treatment for, 646,  
648–656  
toxicological dose- or exposure-  
response relationship in, 109
- Organs, target  
xenobiotic uptake by, 78
- Oropharynx  
respiratory tract protection role of,  
137
- OSH Act, 1453–1454, 1462, 1464, 1465
- OSHA  
Mine Safety and Health  
Administration (MSHA) separation from, 1464
- Osler, William  
coal worker health problems, studies  
by, 22
- Ottawa Charter, 1222

- Out-migration of workers  
cross-sectional studies affected by,  
236–237  
medical record information  
concerning, 280  
with occupational asthma, 236–237
- Outcomes (in epidemiology)  
categories, assigning of, faulty,  
285–286  
of interest, 259, 260–263  
measures, accuracy of, 205  
overview of, 259  
and risk estimates, 263–271
- Outsourcing, 1541  
employee health programs affected  
by, 36–37
- Owen, Robert (1771–1858)  
worker communities established by,  
1203
- Oxidant gases  
exposure to, 90–91
- Oxides  
of nitrogen, 630–637, 755  
of sulfur, 637–640, 636*f*, 637*t*, 755,  
757
- Oxon pesticide group. *See* Carbamate  
pesticide
- Oxygen deficiency  
loss of consciousness due to, 1110
- Ozone  
depletion of, 800–801  
economic impact of problems with,  
757  
mortality associated with, 752  
as pollutant category, 755  
production of, 746  
respiratory symptoms affected by, 753  
toxic gases, comparison to, 141
- Pacific Presbyterian Medical Center  
(San Francisco), 1267
- Padua, 18
- Pain  
conceptual model of, 1001*f*  
perception and nature of, 930–931,  
970–971, 1001–1006  
workers' compensation claims for,  
1335–1336
- Pan American Health Organization  
(PAHO), 1478, 1578
- Pancreas  
ethanol impact on, 122
- Pandemic preparedness, 1380–1384
- PAPS, 90
- Paracelsus (Philippus Aureolus Bom-  
bastus von Hohenheim) (1493–  
1541), 16  
toxicology, contributions to, 60–61  
toxicology, observations in, 63, 106
- Parathion pesticide  
toxicity of, 649
- Parkinson's disease  
conditions resembling, 1115, 1114*f*  
pesticides as risk factor for, 228
- Part-time personnel  
in occupational health services, 1283
- Partial absence (defined), 1164
- Partial impairment  
rating, 1181–1182
- "Partially fit" (term), 1147
- Particles  
airway, behavior in, 131*f*  
deposition of, 124*f*, 130–133  
excretion of, 98  
in inhalation toxicology, 144  
penetration and deposition of, 126  
sedimentation velocity of, 132  
size of, 128–130  
types of, 127–128
- Particulate air pollution  
exposure to, 136  
lung cancer risk association of, 1090

- Particulate matter  
recognition of, 755  
types of, 756
- Patch testing  
for skin disorder evaluation, 1025–1026
- Patient-centered care *versus* client-enterprise-provider relationship, 1262
- PCBs (polychlorinated biphenyls),  
596f  
bans on, 799  
child exposure to, 99  
overview and health effects of, 665–667  
reproductive problems caused by, 174  
thyroid hormone activity affected by, 175–176
- Pediatric environmental health, 801–807, 805, 1450
- Pediatric Environmental Health Specialty Units (PEHSUs), 806–807, 1020
- Pelvis  
autonomic dysfunction of, 974–975
- Pension system  
in Germany, 26–27
- Pentachlorophenol  
molecular structure of, 596f, 661  
use and toxicity of, 661
- People, a (defined), 1542
- Perchlorate  
water contamination with, 767
- Perfluorooctanoic acid (PFOA) (C8)  
water contamination with, 767
- Performance-related evaluations  
overview of, 1142, 1148–1152
- Period prevalence  
defined, 254
- Peripheral nerves  
injuries to, 999
- Peripheral neuropathy  
toxic, 1117–1119
- Permissible Exposure Limits (PELs)  
for gases, 135
- Persistent toxic substances (PTSs)  
overview of, 789–790
- Personal competence  
as mental health component, 819
- Personal-exposure monitoring  
overview of, 323
- Personal injury cases, 1390
- Personal injury law, 1393  
*See also* Tort litigation
- Personal leave  
absence due to, 1163  
managing, 1165–1166  
types of, 1162t
- Personal problems  
employee assistance programs  
addressing, 1233, 1234
- Personal protection  
eye protection, 873–875  
hearing protection, 862–865  
as occupational hazard control strategy, 849–850  
overview of, 851–852  
protective equipment (*see* Protective clothing; Protective equipment)  
respiratory protection, 853–862  
skin protection, 865–873
- Personal risk factors. *See*  
Individualized risk factors
- Personal sampling  
as exposure assessment tool, 320
- Personality characteristics  
as burnout factor, 834–835
- Personnel issues  
occupational health service involvement in, pitfalls of, 1274
- Personnel records  
in case-referent studies, 251

- Personnel records (*continued*)  
in epidemiology studies, 272, 281–283  
in prospective studies, 249
- Pesticides  
chemical classes of, 645t  
definition and overview of, 640–643, 646  
environmental medicine cases involving, 41  
human consumption of, 169  
infertility and reproductive problems caused by, 172, 174  
obsolete, 797645t  
as Parkinson’s disease risk factor, 228  
regulation of, 1451  
testing of, 188, 190  
toxicity, derivation of, 649  
toxicity, managing, 642t–643t  
transcutaneous absorption of, 77  
*See also specific pesticide*
- Petersen, Dan  
safety theories developed by, 366
- Phalen’s test for carpal tunnel syndrome, 998f
- Pharmaceutical agents  
in surface water, 766–767
- Pharmaceuticals  
testing, 190
- Pharmacodynamics  
defined, 55
- Pharmacokinetics  
defined, 55
- Pharmacology  
toxicology as subdiscipline of, 61
- Phased return to work  
overview of, 955–956
- Phenoxyacetic herbicides  
use and health effects of, 662–664
- Phocomelia  
definition and overview of, 180
- Phosphoric acid  
as water additive, 765–766
- Photochemical air pollution  
identification of, 751  
overview of, 745, 746–747
- Physical, passing  
as hiring requirement, 1144
- Physical conditioning  
as health promotion component, 1222 (*see also* Fitness programs)
- Physical labor  
ability, screening employees for, 1145
- Physical therapy  
for injured workers, 956–957
- Physicians  
as environmental advocates, 46–49  
exposure assessment role of, 295  
healthcare plans, attitude concerning, 30, 31  
medieval masks worn by, 15–16, 15i  
non-specialist in occupational medicine, 5
- Physicians’ assessments  
precautions concerning, 279
- “Physician’s first report”  
legal requirements for, 1331–1332
- Physicians for Global Survival (Canada), 48
- Physicians for Social Responsibility, 48
- Physiologically based pharmacokinetic models (PBPK)  
overview of, 104–105
- Physiology  
toxicology as subdiscipline of, 61
- Pipefitters  
cancer risks among, 1091
- Plague  
protective measures against, 15–16
- Planning and intelligence  
in incident command system (ICS), 1377–1378

- Plant engineering records and histories in epidemiology research, 272, 283
- Plant pesticides  
human adaptation to, 169
- Plant physicians  
duties and authority of, 1251, 1252
- Plants  
studies based at specific, 234
- Pleura  
asbestos effects on, 575  
chemical carcinogens affecting, 153  
plaques, 1055–1056, 1054*f*, 1056*f*
- Pneumoconioses  
overview of, 1036–1039  
studies on, 22  
*See also specific pneumoconiosis type*,  
e.g.: Asbestosis
- Pneumonia  
alcoholism associated with risk of, 142  
as building-related disease, 694–695
- Point prevalence  
defined, 254
- Point-source emissions  
overview of, 747
- Poisons  
historic use and antidotes for, 58,  
59–60
- Police, fitness for duty evaluations for,  
1153
- Pollutants, fate and disposition of, 798
- Pollution, regulation and data collecting on, 772–775
- Pollution exposure, socioeconomic status, low linked to, 225
- Polybrominated biphenyl ethers (PBDEs), 768
- Polychlorinated biphenyls. *See PCBs* (polychlorinated biphenyls)
- Polycyclic aromatic hydrocarbons (PAHs)  
as carcinogens, 152, 1094–1095
- overview and health effects of, 667–668
- Polydispersed aerosols  
overview of, 128
- Polyethylene garnets, 873
- Polyethylene gloves, 871
- Polymer fume fever  
overview of, 1085
- Polyurethane gloves, 871
- Polyvinyl chloride (PVC) gloves, 871
- Pontiac disease  
as building-related disease, 694–695
- Pooling  
limitations of, 223  
overview of, 223
- Popper, Karl  
scientific theory falsifiability illustrated by, 211
- Population  
studies, 1424  
workforce relationship to general, 341*f*
- Population-based monitoring  
historically important studies, 343*t*  
occupational health trends, surveillance for, 348–353  
overview of, 348  
sentinel events, 353–355
- Population-based screening programs  
overview of, 500–501
- Population health  
definition and term usage, 1219–1220  
goals in, 1221–1222  
historic concepts of, 1205  
improving of general, 237  
model of, 1217*f*  
*versus* personal health, 1546–1547
- Population health framework  
overview of, 486–489
- Population-level interventions  
as preventive medicine component, 1217

- Population medicine  
definition and term usage, 1217, 1219–1220
- Positive stress  
overview of, 812–813
- Possible (defined), 1413
- Post hoc, ergo propter hoc* (concept), 1394
- Posttraumatic stress disorder  
causes of, 817
- Potential energy  
definition and components of, 505–506
- Pott, Percivall  
chimney sweeps, cancer among noted by, 64, 154
- Power  
curves, 219f  
definition and overview of, 218–223
- Pralidoxime (2-PAM, 2-pyridine-aldoxime methiodide)  
organophosphate and carbamate pesticides, 651–656
- Pre-contemplation (as behavioral change stage)  
overview of, 494–495
- Pre-employment examinations  
faulty, examples of, 1136t  
overview of, 890–891  
pre-placement evaluations  
compared to, 1143–1146  
term, 327
- Pre-existing conditions  
exacerbation *versus* aggravation of, 1176–1177
- Pre-placement drug testing, 1190
- Pre-placement evaluations  
essential features of, 1138t  
job change-initiated evaluations  
compared to, 1152  
low-back pain risks, determining, 985–986
- medical records as source of, 277
- occupational health center involvement in, 1268
- overview of, 890, 1142, 1143–1146
- return-to-work evaluations compared to, 1148
- Precautionary principle  
in risk management, 431
- PRECEDE/PROCEED model  
overview of, 492–493
- Predictive toxicology  
terminology of, 184t–185t
- Pregnancy  
ergonomic issues during, 1106  
occupational safety issues during, 1104–1105  
women's absence for reasons other than, 1165  
work absence due to, 1163
- Prepaid care  
roots of, 1205
- Prepaid health services  
advocacy of, 30
- Prepaid healthcare  
attitudes concerning, 30
- Prepaid medical care plan  
origin/model of, 25
- Prepaid practice plans  
rise of, 31
- Preparation (as behavioral change stage)  
overview of, 496
- Preponderance of evidence, 1418, 1417t, 1420–1421
- Presenteeism  
definition and decreasing, 1239
- Pressure  
as skin disorder cause, 1030
- Pressure and altitude  
overview of, 544–550

- Presumption  
 legal application of, 1426–1427,  
 1428–1432  
 in occupational disease, 1347,  
 1348–1350
- Prevalence  
 defined, 335
- Prevalence data  
 cross-sectional studies providing, 254
- “Preventative” (term), 1217
- Prevention science  
 application of, 1201  
 basics of, 462  
 definition and overview of, 459  
 prevention approaches and strategies,  
 474–477, 475f, 1206, 1207
- Preventive medicine  
 commitment to and attitudes con-  
 cerning, 460–461  
 as health promotion component,  
 1223–1224  
 health services guidelines, 1212,  
 1213–1212 (physician  
 compliance)  
 historic context, 1202–1203, 1207  
 levels of, 463–471, 464f  
 nature and scope of, 47, 459–460,  
 462, 1217  
 occupational and environmental  
 medicine as speciality of, 1201  
 occupational medicine services,  
 1248–1249  
 opportunities, expanded in, 49  
 outsourcing impact on, 37  
 prevention, approach to, 474–475,  
 476–477  
 technology of, 461–462  
 terminology of, 1216–1221  
 in workplace, 462
- Preventive medicine services  
 reimbursement for, 1226
- in service industries, 1280
- Preventive medicine specialists  
 environmental health issues,  
 involvement in, 38–39, 740  
 focus, areas of, 7–8  
 “Preventive” (term), 1217
- Previous jobs  
 health problems stemming from, 237
- Primary care  
 company on-site clinics for, 38  
 in developing countries, 1553  
 occupational medicine in, 1260,  
 1261–1264, 1279
- Primary prevention  
 overview of, 462, 463, 465–469  
 public health triad in, 471
- Primary wastewater treatment  
 overview of, 760
- Private sector  
 safety-sensitive jobs in, 1154–1155
- Probabilities  
 balance of, 1412, 1413–1414, 1416  
 expressing, 270–271
- Probable (defined), 1413
- Productivity  
 absence effect on, 1161–1162  
 assessment instruments, 1239–1240
- Productivity costs  
 conditions linked to greatest, 1237
- Productivity management  
 wellness and, 1238–1240
- Productivity-work hours utility curve,  
 366
- Professional liability insurance, 1438
- Program evaluation  
 overview of, 259
- Progression period  
 in cancer, 156
- Progressive Era  
 occupational medicine during, 12,  
 13, 26–29

- Progressive Era (*continued*)  
workers' compensation during, 1326, 1327
- Progressive massive fibrosis  
overview of, 1060–1061
- Prohibition  
alcoholic poisoning during, 62–63
- Proliferation phase (in cancer)  
overview of, 156, 161
- ProMed-Mail  
travel medicine information  
provided through, 1571
- Promotion phase (in cancer)  
overview of, 156, 160  
threshold of, 156
- Proportionate analysis  
overview of, 270
- Proportionate mortality (or morbidity) study  
overview of, 249
- Proportionate mortality ratio (PMR)  
risk, expressing as, 268, 270
- Propylene glycol  
toxicity of, 571
- Prospective (cohort) studies  
biases in, 235–238 (alt. term: cohort studies)  
case-referent studies compared to, 256, 271  
overview of, 244–249
- Prostate cancer  
detection of, 162–163
- Protection. *See* Personal protection
- Protective clothing  
in cold weather, 556–558  
footwear, 875–876  
gloves, 868–872  
overview of, 872–873
- Protective equipment  
for eyes, 873–875  
gloves, 868–872  
overview of, 850t
- respirators, 851t, 854–862
- Proximate cause  
defined, 1405–1406
- Pseudomonas* outbreaks  
sources of, 763
- Psychogenic outbreaks  
anatomy of, 726–734  
dynamics of, 722–725  
epidemic curve, 732, 730f  
features suggesting, 721t  
features unreliable in suggesting, 722t  
identifying, 692, 709–710  
overview of, 721–722  
possibility of, 718  
propagation trees for, 731f  
reconstruction of, 727t, 729t  
typical scenario of, 725–726
- Psychological evaluations  
outside employee assistance programs, 1236
- Psychosocial stress  
cross-sectional study application to, 253
- Psychosomatic illnesses  
workplace stress as factor in, 832
- Public, communication with  
practical skills for, 419–422
- Public accommodation  
barrier-free access in, 1135
- Public health  
definition and term usage, 1217, 1218–1219, 1220–1221  
environmental health and medicine relationship to, 7, 739t  
environmental issues and, 739  
exposure control, 1202  
historic regulations concerning, 23–24, 62  
as occupational and environmental medicine component, 1481, 1482  
reforms in, 485–486  
risk science applied to, 379–380

- safety and injury prevention,  
approaches to, 368–369  
workplace, protection in, 39
- Public Health Act of Parliament,  
Great Britain, 1848, 23, 62
- Public Health Agency of Canada  
travel medicine information  
provided by, 1571
- Public health approach to prevention  
overview of, 474, 475–476, 477
- Public health model  
overview of, 43–44
- Public health specialists  
environmental health issues,  
involvement in, 38–39, 740
- Public health standard  
in risk management, 432
- Public health triad  
applications of, 472*f*, 473*f*  
overview and components of,  
471–474
- Public safety occupations  
psychological adjust factors in, 1373
- Publication bias  
risk detection limited by, 223
- Pullman, George  
worker health, approach to,  
1203–1204
- Pulmonary edema  
management of, 1067  
toxic gases as cause of, 134
- Pulmonary function tests  
smoking information covered in, 279
- Pyrethroids  
use and toxicity of, 661–662
- “Qualified individual”  
defined, 1134
- Qualitative risk assessment  
definition and overview of,  
384–385  
limitations of, 399–403
- Quality of working life, 1495–1496
- Quantitative fit testing, 860–861
- Quasi-experimental evaluation study  
designs  
overview of, 258
- Radiation  
exposure, monitoring, 319–323, 320  
regulations concerning, 529–531  
*See also* Ionizing radiation; Nonionizing radiation
- Radiation-induced cancer  
risk of, 525–528  
susceptibility to, 163
- Radiofrequencies below microwave  
overview of, 539–540
- Radioimmunoassay  
as drug testing technique, 1194–1195
- Radiological agents, terrorist use of,  
1386–1387, 1386–1387
- Radon  
environments where found, 698–700
- Railroad workers  
workers’ compensation for, 1322
- Railroads  
medicine and surgery, historic role  
in, 24–26, 1247
- Ramazzini, Bernardino, 17*i*  
legacy of, 22  
life and career/occupational medicine,  
contributions in, 16–19  
occupational cancer observed by, 154  
toxicology, observations in, 63
- Random error  
effect of, 224  
overview of, 216–217
- Random events  
clusters and, 239
- Rantanen, Jorma, 33
- Rare diseases and disorders  
case disease and report application  
to, 257–258

- Rare diseases and disorders (*continued*)  
studying, challenges of, 285–286
- Rats  
cancer in, 182, 191  
life span of, 189  
toxicology experiments in, 118
- REACH (Registration, Evaluation, and Authorization of Chemicals)  
overview of, 190–191
- Toxic Substances Control Act (TOSCA) replaced by, 773
- Reactive airway dysfunction syndrome (RADS)  
overview of, 1077–1080
- Reactive airway disease  
building-associated triggers of, 700
- Reagan, Ronald, 1456
- “Reasonable accommodation” (term), 1133–1134, 1185, 1187
- “Reassignment to a vacant position”  
defined, 1186
- Rebuttable presumption  
legal application of, 1428, 1429  
in occupational disease, 1348–1349, 1407
- Receptors  
xenobiotic concentration at level of, 78
- Records in epidemiology  
death certificates, 204, 272  
exposure records, 272–273, 276  
medical records, employees’, 272, 276–281  
personnel records, 272, 281–283  
plant engineering records, 272, 283
- Recovery  
goals and expectations for, 944, 954–956, 955–956
- Recovery, incomplete  
absence due to, 1167
- Recreational waters  
quality of, 763
- Rectal cancer  
risk factors for, 123
- Recycling  
overview of, 771
- Reducing systems  
overview and examples of, 94–96
- Reference dose (RfD)  
defined, 443–444
- Reference group or population  
definition and function of, 246  
overview and function of, 265–266  
relative risk calculated using, 266
- Reference population  
function of, 247
- Referral  
occupational health services role in, 1279, 1282  
overview of, 913–915
- Regional complex pain syndrome  
workers’ compensation claims for, 1336
- Regression-based approaches  
ecological association study  
methods compared to, 255–256
- Regression techniques  
conventional analysis methods  
replaced by, 256  
in perspective studies, 247
- Regulatory toxicology  
benchmarks in, 187*f*  
overview of, 180–182
- Rehabilitation Act, 1134–1135
- Rehabilitation counselors  
duties of, 1254
- Relative risk  
calculating and expressing, 266–267  
prospective study showing, 247–248  
and related measures, 265–270, 271  
as standard mortality ratio (SMR), 248
- Relativism, 1508, 1509
- Repetitive strain injury  
cervical syndrome, 992–993

- conditions and diagnostic features of, 989t  
overview of, 988–990, 991t, 992  
tendonitis, tenosynovitis, bursitis, 995–996  
tension neck syndrome (tension myalgia), 992  
thoracic outlet syndrome, 993–995  
in workers' compensation, 949
- Reproduction and development  
occupational hazards to, 1103–1106
- Reproductive toxicology  
developmental toxicology compared to, 177  
endocrine disruption, 174–176  
overview and scope of, 171, 172–174
- Residential areas  
occupational clinics located near, 1280–1281
- Resilience  
as mental health component, 819  
as stress coping tool, 811–812
- Resistance (of toxic effects)  
defined, 115
- Resource Conservation and Recovery Act, 1976 (RCRA), 774, 1448
- Respirators  
fit testing, 858–862  
types of, 851t, 854–857
- Respiratory carcinogenesis  
particles producing, 126–127
- Respiratory diseases  
children's, nitrogen dioxide associated with, 142  
clinical assessment of, 1035–1036  
overview of, 1030–1032  
*See also specific respiratory disease, e.g.:*  
Lung cancer
- Respiratory excretion  
overview of, 98
- Respiratory protection  
job change-initiated evaluations for jobs requiring, 1152–1153  
overview of, 853–862
- Respiratory range  
defined, 133
- Respiratory tract  
defense mechanisms of, 136–143, 138t–139t  
disorders of, 1029t  
particle behavior in, 130–133  
*See also Lung*
- Responsible Care®, 779–780, 1380
- Restaurant inspection  
food protection through, 795
- Retired employees  
records of, 281–282
- Retirement  
diseases following, 282  
healthy worker effect persisting beyond, 236  
occupational epidemiology study cutuff point at, drawbacks of, 248–249
- Retrospective studies. *See Case-referent studies*
- Return-to-work evaluations  
employee assistance programs and, 1235  
exposure circumstance investigations compared to, 1020  
functional assessment as component of, 1019  
medical records as source of, 277  
modified duty following, 1185  
outcome of, 947  
overview of, 891, 1142, 1146–1148  
readiness and timing of, 945  
in workers' compensation cases, 1349
- Return-to-work goals and  
expectations  
overview of, 955–956, 957–958

- Rhabdomyolysis  
agents and trauma causing, 116
- Rheumatoid arthritis  
coal workers' pneumoconiosis associated with, 1063
- Risk  
defined, 379, 383, 877  
detecting excess, 221*t*, 222
- Risk analysis  
definition and overview of, 381, 382–383  
as risk evaluation step, 380
- Risk assessment  
assumptions in, 394*t*  
classifying and prioritizing, 881–882, 882*t*  
defined, 876  
definition and overview of, 381, 384–388  
hazard identification and characterization in, 388–389, 392  
limitations of, 399–403  
models, 402*f*  
paradigms, 380*f*, 382*f*, 384*f*  
toxicology application to, 180–181  
*See also* Enterprise risk assessment; Qualitative risk assessment
- Risk-benefit analysis  
in risk management, 437
- Risk characterization  
defined, 380  
definition and overview of, 392–393, 395–403, 407*t*
- Risk communication  
approaches, alternative to, 416–418  
community right to know, 429–430  
definition and overview of, 381, 410–412  
as environmental medicine component, 39  
public and media appearances, practical skills for, 418–422  
seven cardinal rules of, 412–416, 412*t*  
worker education and right to know, 422, 424–426  
worker notification, 426–429  
written communication, 422
- Risk comparison  
definition and overview of, 381
- Risk estimates, 263–271  
comparison for occupational cohort, 269*t*  
misclassification effect on, 224  
for prospective studies, 247–248
- Risk evaluation  
approach to, 379  
steps of, 380–381
- Risk factors  
causation referring to, 230  
measurement, availability of, 228  
multiple, 229, 367–368  
non-work related, 285  
reducing, 462  
*See also* Individualized risk factors
- Risk management  
action taking, criteria for, 430–433  
approaches to, 379, 434*t*  
concept, development of, 367  
definition and overview of, 381, 382, 430, 876  
emergency management *versus*, 1369  
near misses, reporting role in, 369–370  
options, overview and assessment of, 433–439, 439*t*  
paradigms, 380*f*, 382*f*  
social dimensions of, 397*f*  
standards and guidelines in, 439–440, 441–442  
terminology for, 443*t*–444*t*  
through business life cycle, 878–880, 879*t*
- Risk perception  
definition and overview of, 381, 403–410

- physician influence on, 47  
risk acceptance *versus* intolerance, 406–408, 407t
- Risk profile, building  
in workers' compensation, 1344–1345
- Risk reduction and control  
overview of, 878
- Risk regulations  
standards and guidelines in, 439–440
- Risk-risk analysis  
in risk management, 437
- Risk science  
definition and scope of, 379–382  
terminology of, 390t–391t
- Risk-taking behavior  
discouraging, 1239  
positive stress in form of, 813
- RNA  
toxin effect on, 87
- Rodent carcinogens  
health hazards, human determined  
on basis of, 169–170
- Rodent studies  
in stress response, 814, 815
- Rom, William  
works of, 44
- Romans, ancient  
chemical hazards, regulation by, 58–59  
occupational medicine of, 14
- Roosevelt, Franklin D.  
mobilization undertaken by, 29
- Rose, Sir Geoffrey, 475–476
- Rose paradox, 476
- Rotating shift  
pros and cons of, 830
- Rothman, Kenneth  
methodological research by, 256  
writings of, 231–232
- Rounded atelectasis (folded lung)  
overview of, 1057, 1055f
- Rounded lung syndrome  
overview of, 1056–1057
- Royal College of Physicians and Surgeons, 9
- Rubber gloves, 871–872
- "Rule of halves," 1012–1013, 1012f
- Rural groups  
health promotion program  
influence limited for, 1228
- Rural workers  
health promotion program  
influence limited for, 1227
- Rush, Benjamin  
influence of, 21–22
- Sackett, David, 920  
bias sources compiled by, 224
- Sacral roots  
compression of, 974
- Sacroiliitis  
as low-back pain factor, 976
- Safety  
in occupational context, 1277, 1278–1279, 1276t  
organizational responsibility for, 1277
- Safety committee minutes  
for incident investigation, 277
- Safety engineering  
emergency management and, 1369  
occupational injury prevention role of, 951
- Safety engineers  
duties of, 1251, 1254  
hazard control role of, 845  
job description and duties of, 847
- Safety factor (SF)  
defined, 444
- Safety practices  
overview of, 369–370, 372–373  
programs and standards, 374–377, 376t, 445, 446f, 447–448

## Index

- Safety science  
medical applications of, 357  
multiple factors, seeking in, 367–368
- Safety-sensitive positions  
drug screening for, 1188, 1189  
evaluations for, 1142, 1153–1155
- Saliva  
excretion role of, 99
- Salmonella*, 791, 792
- Sanitary landfills  
fires at, 772  
site selection for, 768  
as waste disposal method, 770–771
- Sanitary Revolution  
overview of, 23
- Sanitary Revolution (United Kingdom), 485
- SARS  
outbreak of 2003, 1360–1361, 1381  
overview of, 790–791  
principles applicable, 1383  
workplace risks of, 1122
- Scalds  
prevalence and occupational association of, 968
- Schedule  
in occupational disease, 1347–1348, 1407
- Schilling, Richard  
life and work of, 34
- Schistosomiasis  
sources of, 763
- Schools  
building-associated outbreaks in, 689
- Schwarzenegger, Arnold, 1329
- Sciatica  
as low-back pain factor, 977  
subjective nature of, 980–981
- Science  
strengths and limitations of, 210–211
- Scientific evidence *versus* weight of evidence, 1414–1415
- Scientific standard  
in risk management, 432–433
- Screening occupational and environmental history  
overview of, 895–896
- Screening tests, medical  
Americans with Disabilities Act  
impact on, 1145  
definition and purpose of, 327, 1210–1211  
executive health programs  
combined with, 1216  
history of, 327–328  
performance of, 335–340, 336*t*  
selection of, 339*t*
- Seamen  
workers' compensation for, 1322
- Sears Roebuck & Co., 1205
- Seasonal industries  
occupational health services in, 1283
- Second injuries  
defined, 1182
- “Second injury funds”  
definition and overview of, 1340
- Second-order kinetics  
conditions resulting in, 102–103  
overview of, 104
- Secondary prevention  
overview of, 462, 463, 469–471
- Secondary wastewater treatment  
overview of, 760–761
- Segmental vibration  
effects and disorders, 566–568  
exposure, extent of, 564
- Self-blame  
effects of, 926  
incident cause theories based on, 363, 364

- Self-contained breathing apparatus (SCBA)  
toxic gas protection with, 605
- Self-referral  
to employee assistance programs, 1233, 1234
- Selikoff, Irving J. (1915–1992)  
asbestos studies by, 582  
influence of, 13  
work and writings of, 35–36
- Selye, Hans, 815
- Sensitizer-induced occupational asthma  
overview of, 1072–1076  
sensitizing agents, 1071<sup>t</sup>
- Sensory neuropathies  
overview of, 1117
- Sentinel events  
in population-based monitoring, 353–355
- “Serious health problem” (defined), 1157
- Settlement movement  
overview of, 1204
- Sex ratio  
environmental factors affecting, 176
- Sexually transmitted infections  
among sex workers, 1124  
among travelers, 1572
- Shewhart, Walter A., 877
- Shift work  
stress and health effects of, 825–831
- Sick building syndrome (SBS)  
assessment of, 706–709  
causes of, 721  
definition and overview of, 691  
environmental medicine cases stemming from, 40  
low-grade symptoms other than, 695  
overview of, 704–706
- Sick leave  
abuse of, 1167  
family and medical leave concurrent with, 1157  
policy compliance, enforcement of, 1160
- Sickness absence  
chronic illness-related, 1162  
defined, 1163  
prevalence of, 1159  
types of, 1162<sup>t</sup>
- Sidestream cigarette smoke  
constituents of, 715<sup>t</sup>  
defined, 716  
persons with atopy affected by, 702
- Siegrist, Johannes, 814
- Silent Spring* (Carson), 660
- Silica  
carcinogenic properties of, 165  
overview and health effects of, 668–670
- Silicates  
fibrous (*see* Fibrous silicates)  
sources of, 575  
*See also Type of silicate, e.g.:*  
Asbestos
- Silicosis and silica-related disease  
comorbidities, 1047, 1049–1050, 1047<sup>f</sup>, 1048<sup>f</sup>  
as hilar lymphadenopathy, 1042<sup>f</sup>  
overview of, 1041–1050
- Silicotic nodule  
appearance of, 1038<sup>f</sup>  
description of, 1042
- Silicotuberculosis  
appearance of, 1046<sup>f</sup>
- Silo filler’s disease  
appearance of, 1063<sup>f</sup>  
causes and prevention of, 632, 1064
- Silver nanoparticles  
water contamination with, 768

- Simple coal workers' pneumoconiosis  
    appearance of, 1059f
- Simple silicosis  
    appearance of, 1041f, 1043f  
    lung cancer accompanying, 1048f  
    overview of, 1042–1044
- Simulation  
    exposure estimate based on, 315–316
- Sinclair, Upton  
    writings of, 62
- Single-employer community-based healthcare services  
    overview of, 1257, 1259
- Single-employer internal services  
    overview of, 1257, 1258–1259
- Skin  
    chemical carcinogens affecting, 153  
    xenobiotic absorption through, 70, 72, 77
- Skin cancer  
    as occupational cancer, 1095, 1097  
    risk factors for, 64
- Skin disorders  
    acute skin injury, 1022–1024  
    overview of, 1020–1022  
    prevalence of, 1015  
    types of, 1024–1030
- Skin protection  
    contaminated clothing, handling, 866–867  
    hand protection, 867–872  
    overview of, 865–866
- Small businesses  
    health services for, 1305–1310
- Smoke  
    content of, 140  
    exposure to, 126
- Smokers  
    absenteeism among, 1165
- Smoking  
    air pollution effects compared to, 141  
    alcohol consumption associated with, 226  
    asbestos effects combined with, 1055, 1098f, 1405–1406  
    gastrointestinal cancers linked to, 123  
    inhalation toxicology and, 124  
    as lung cancer risk factor, 276, 1098f  
    lung carcinogen interaction with, 1096  
    medical record information concerning, 276, 278, 279  
    respiratory tract affected by, 140–141  
    social attitudes concerning, 1224  
    socioeconomic status, low linked to, 225, 227
- Smoking cessation  
    as health promotion program component, 1227, 1231
- Smoking-related problems  
    productivity costs and expenses of, 1237  
    “So far as is reasonably practicable” (SFAIRP)  
        as risk management approach, 434–435
- Social capacity  
    protection, wellness program role in, 1238
- Social Darwinism  
    safety approach influenced by, 363
- Social epidemiology  
    update to approach of, 226
- Social insurance  
    in Germany, 26–27
- Social marketing  
    definition and overview of, 1206–1207
- Social marketing theory  
    overview of, 493–494

- Social science  
medicine as, 1484–1485
- Social Security disability insurance  
overview of, 1178
- versus* workers' compensation, 1340, 1341, 1352
- Social support  
components and effects of, 819  
as stress coping tool, 811
- Socialized medicine  
attitudes toward, 30
- Socially isolated groups  
health promotion program  
influence limited for, 1228
- Socioeconomic status, low  
health risks associated with, 225–226
- Socrates, 59
- Soil contamination  
biotransformation in, 797
- Solid and municipal waste  
composition of, 769, 767*t*  
disposal of, 768–772, 781–782
- Solubility  
as gas toxicological behavior factor, 133–134
- Soluble gases  
health effects of, 134
- Soluble xenobiotics  
inhalation of, 74
- Solvents  
chemical classes of, 670*t*  
ethyl alcohol compared to, 570  
health effects of, 1111, 1115  
neurological effects of, 1118  
overview and toxicity of, 570, 671–776  
reproductive problems caused by, 174  
skin contact, effect of, 1025*f*  
transcutaneous absorption of, 77
- Somatization  
conversion reaction compared to, 934, 935  
definition and overview of, 931–934  
as low-back pain factor, 977
- Somatization disorder  
somatization compared to, 932
- Somers, Anne, 1212
- Sound level meters (defined), 541
- Specialty consultation  
occupational health services role in, 1279, 1282
- Speciation  
defined, 681
- Species  
interspecies differences, 170  
strain differences within, 182
- Specificity  
as causation criterion, 214
- Spheroidal particles  
overview of, 127
- Spinal stenosis  
as low-back pain factor, 974, 976*t*  
workers with, identifying, 986
- Spiral CT scanning  
evaluation of, 162
- Sports medicine  
functional capacity evaluation in, 1170
- “Squeeze”  
overview of, 546
- Standard of persuasion and belief, 1418, 1417*t*, 1421–1423
- Standardized mortality ratio (SMR)  
defined, 212–213  
likelihood assessment, Hill criteria for, 212–216  
relative risk, presenting through, 267–268, 270  
relative risk as, 248
- Standards  
defined, 439  
setting, 440–442, 443*t*–444*t*  
terminology for, 443*t*–444*t*  
types of, relationships of, 446*f*

- Standby emergency treatment (SBET)  
for malaria, 1574
- State-level drug testing regulations  
in transportation industry, 1197
- Statistical analysis  
random error possibility in, 216–217
- Statistical quality monitoring theory, 877
- Stockholm Convention on Persistent  
Organic Pollutants (“POPs  
Treaty”), 789–790, 799
- Stomach  
toxic metal impact on, 122
- Strength of evidence, 1417t
- Stress  
absence due to, 1167  
behavioral response to, 817  
body’s response to, 814–816  
clinical responses to, 817  
human *versus* animal response to,  
814–816  
individual response to, 817–821  
job-related, 810–811  
positive forms of, 812–813  
response, musculoskeletal disorders  
and, 970–971  
sources of, 810, 811
- Stress in workplace  
gender-related issues, 820  
health effects of, 831–832  
overview of, 809–814  
personal life stress combined with,  
819, 820  
psychological effects of, 811–812, 832  
reduction and management of,  
837–842, 836t–837t, 839t  
shift work, 825–831  
sources of, 810–811, 821, 820t–821t  
work organization, 821–825
- Stress management  
as health promotion program com-  
ponent, 1232
- Stressors  
definition and overview of, 811
- Strict Father *versus* Nurturing Parent  
concept, 1501–1507
- Strychnine  
absorption of, 75
- Study designs, inferential  
aspects of, 256–259  
case-referent studies, 250–252  
cross-sectional studies, 252–254  
ecological association studies,  
254–256  
overview of, 244  
prospective studies, 244–249  
strengths, relative of, 257f
- Study findings  
detecting, probability of,  
218–219
- Study subjects  
identification, accuracy of, 234
- Styrene  
elimination of, 104
- Subacute combined system atrophy  
overview of, 1115
- Subclinical findings  
identifying, 260
- Subcontractors’ employees  
documentation for, 282
- Subgroups of workers  
identifying, 280
- Substance abuse  
chronic pain linked to, 1005  
identifying, 1150, 1167  
productivity costs and expenses of,  
1237  
programs addressing, 1232
- Substantial causation  
overview of, 1411–1412
- Substitution and elimination  
as occupational hazard control  
strategy, 848–849

- Suicide attempts  
  ingestion in, 70, 75, 76
- Sulfotransferase  
  chemical structure of, 89*f*  
  function and action mechanisms of, 90  
  metabolism of, 90
- Sulfur  
  oxides of, 637–640, 636*f*, 637*t*, 755, 757
- Sulfur dioxide  
  exposure to, 136  
  as reducing air pollution component, 744  
  toxic gases, comparison to, 141
- Sulfur dioxide–particulate complex  
  overview of, 136
- Superfund, 1449
- Superfund Amendments and Reauthorization Act, 1986 (SARA), 429, 786–787, 788
- Superoxide dismutase (SOD)  
  antioxidant role of, 92–93
- Superoxides  
  formation and reduction of, 92*f*
- Supervisors  
  employee assistance programs, role in, 1233, 1235
- Supraspinatus tendonitis (rotator cuff tendonitis)  
  overview of, 996
- Surface water  
  contamination of, 762  
  movement and quality of, 757–758  
  pharmaceutical agents in, 766–767
- Surrogate measurements  
  exposure estimate with the aid of, 316–317
- Surveillance  
  baseline for, 252–253  
  defined, 327, 891
- See also* Monitoring and surveillance
- Surveys  
  cross-sectional studies in form of, 253, 254  
  reference groups in, 266
- Survival curve  
  linear, 482  
  rectangularization of, 483, 484*f*
- “Survival of the fittest”  
  workers, attitudes toward influenced by, 1203
- Survivor effect  
  mortality and disease risk affected by, 237
- Susan Harwood grants, 1456
- Susceptibility  
  biomarkers of, 262, 286  
  examining, 286  
  to exposure effects, 322–323  
  of person to injury, 358
- Susceptibility states  
  overview and types of, 114–115
- Sustainable development  
  emphasis on, 44
- Sweden  
  occupational healthcare in, 31–32
- Synthetic chemicals  
  *versus* natural pesticides, 169
- Synthetic rubber gloves, 871–872
- Systemic sclerosis (Erasmus’ syndrome)  
  silicosis associated with, 1049–1050, 1047*f*
- Systemic toxicity  
  acute *versus* chronic, 76  
  disorders, prevalence of, 1015  
  kidney involvement in, 118  
  overview of, 1106–1112
- Systems theories of safety and injury  
  prevention  
  development of, 366

- Talc  
asbestos compared to, 578
- Tar burns  
prevalence and occupational association of, 968–969
- Taylor, Frederick Winslow  
injury theories influenced by, 364–365
- TCDD (dibenzodioxin), 597, 600
- TCDF (dibenzofuran), 597, 600
- Technological disasters  
perception of, 403, 404
- Technology  
economic impact of, 1550  
occupational and environmental medicine impacted by, 1481
- Telecommunications Act, 1135
- Telecommuting  
infectious disease spread, preventing through, 1127
- Television viewing time  
autism incidence in studies linked to, 256
- Temperature extremes  
in building environment, 696–698  
overview of, 550–552
- Temperature extremes (cold)  
effects and disorders, 552–556  
protective clothing, 556–558  
respiratory tract affected by, 143
- Temperature extremes (heat stress)  
effects and disorders, 558–564
- Temporal relationship  
as causation criterion, 214
- “Temporary disability” (term), 1176
- “Temporary impairment” (term), 1176
- Tendonitis, 995, 996
- Tennis elbow. *See* Epicondylitides
- Tenosynovitis, 995–996
- Tension neck syndrome (tension myalgia)  
overview of, 992
- TEPP pesticide  
toxicity of, 649
- Teratology  
defined, 177
- Termination (as behavioral change stage)  
overview of, 496–497
- Terrorism  
response to, 1357, 1384–1387
- Terrorist attacks, September 11, 2001, 1359–1360, 1451, 1458
- Tertiary prevention  
overview of, 461, 462, 465, 471
- Tertiary wastewater treatment  
overview of, 760, 761
- Testimony, aids to, 1437
- Tests  
performance, measurements of, 336t
- Tetrachloroethylene  
overview and health effects of, 679
- Tetraethyl lead, 1113
- Texas A&M medical system  
origin of, 26
- Textile industry  
child labor in, 1500
- Thackrah, Charles Turner (1795–1833)  
occupational medicine,  
contributions in, 20, 21*i*  
toxicology, observations in, 63  
writings of, 18, 22
- Thalidomide  
developmental toxic effects of, 179–180
- Therapeutic index  
concept, historic development of, 60
- Thermal insulating gloves, 872
- Thioltransferase  
thiols regenerated through, 91
- Thomas Jefferson University, 1239
- Thoracic outlet syndrome  
Adson’s maneuver for, 992*f*

- overview of, 993–995
- Three Es (safety theory)  
overview of, 367
- Threshold  
in clinical outcomes *versus* toxicology, 110  
defined, 108, 110–111
- Threshold Limit Values (TLVs)  
for gases, 135  
substance exposure detection below, 194
- Thyroid hormone activity  
PCB influence on, 175–176
- Thyroid hormones  
chemicals resembling, 166
- Tight building syndrome (TBS)  
causes of, 713  
definition and overview of, 691–692
- Time series analysis/es  
overview of, 241
- Tin  
organic derivatives of, 1114
- TNT  
use of, 671
- Tobacco smoke  
birth defects caused by, 177  
developmental abnormalities caused by, 178  
reproductive problems caused by, 174
- Tolerance  
concept, historic development of, 60
- Toluene  
overview and health effects of, 679–680
- Toluene diisocyanate, sensitivity to, 1080*f*
- Tort litigation, 1393–1397
- Toxaphene  
as carcinogen, 660
- Toxic disorders. *See* Systemic toxicity
- Toxic effects  
resistance mechanisms, 115, 115*f*
- Toxic encephalopathy  
diagnosis of, 938–939  
overview of, 1112–1115  
questionable diagnosis of, 938
- Toxic exposure  
benchmarks for, 183, 184*t*–185*t*  
respiratory effects of, 1032–1033
- Toxic gases  
air pollutants compared to, 141  
in bloodstream, 134  
diffuse alveolar injury, association with, 1064–1066, 1062*t*  
respiratory tract response to, 1034–1035
- Toxic-induced loss of tolerance  
questionable diagnosis of, 938
- Toxic inhalation  
overview of, 76, 1063–1068
- Toxic mold cases  
prevalence of, 938–939
- Toxic peripheral neuropathy  
overview of, 1117–1119
- Toxic Release Inventory (TRI)  
purpose of, 773
- Toxic Substances Control Act, 1448
- Toxic Substances Control Act (TOSCA), 773
- Toxicity  
action mechanisms of, 105–106  
and carcinogenic activity, correlation between, 168, 170–171
- Toxicity testing  
overview of, 181–182  
types of, 188–191
- Toxicodynamics  
definition and overview of, 55, 105  
exposure–response relationships, 106–115

- Toxicogenomics  
environmental medicine impacted by, 1481–1482
- Toxicokinetics  
definition and overview of, 55, 65–70, 67f  
distribution in, 797  
human *versus* pollutant fate and disposition, 797  
rates and, 100f  
understanding, 100
- Toxicological dose- or exposure-response relationship  
definition and overview of, 108–111
- Toxicological dose-response relationship  
definition and overview of, 106, 108
- Toxicological exposure-response relationships  
carcinogenics and, 168  
early, undetectable phase of, 110–111, 111f  
epidemiological exposure-frequency relationship, translation into, 319–320  
multiple, 109f, 112f  
overview and types of, 106–115, 107f, 108f
- Toxicology  
attitudes concerning, 63  
downsizing and outsourcing of, 36–37  
ecotoxicology compared to human, 797–798  
environmental medicine component of, 40  
in high-technology industries, 1280  
history and concepts of, 58–63  
medical model dominated by, 42
- occupational medicine connection to, 64, 65  
overview and scope of, 56, 57t–58t  
terminology of, 55–56  
*See also toxicology specialty, e.g.: Reproductive toxicology*
- Toxics drop-off facilities  
overview of, 781–782
- Toxics round-up programs  
overview of, 781–782
- Toyota  
in-plant medical services at, 1259
- Trace element imbalance  
questionable diagnosis of, 938
- Trace elements  
definition and overview of, 680–687
- Tracking  
defined, 198–199
- Trade  
globalization and, 1551–1552
- Trades  
personnel records, 282–283
- Trans-theoretical model of behavioral change  
overview of, 494–497
- Transcutaneous absorption  
in occupational medicine, 77
- Transdermal exposure  
action mechanisms of, 72  
as environmental exposure route, 70
- Transmucosal exposure  
mucous membrane absorption, 77  
as workplace exposure route, 70
- Transportation industry  
drug testing for, 1189, 1190, 1196–1197  
injury rates in, 952  
*See also Department of Transportation*
- Transportation workers  
fitness for duty evaluations for, 1153

- Trauma, minor  
  as skin disorder cause, 1030
- Traumatic mononeuropathies  
  overview and types of, 999–1001
- Travel medicine, 1571–1576
- Travelers' diarrhea, 1574
- Tree cutting  
  as high-risk occupation, 960
- Tremolite  
  minerals resembling, 578
- Trench (immersion) foot  
  overview of, 553–554
- Triangle Shirtwaist Company Fire,  
  1911, 27
- Trichloroethylene (TCE)  
  overview and health effects of, 680  
  water contamination with, 764
- 1,1,1—Trichloropropane (TCP)  
  water contamination with, 768
- Trigger finger (stenosing tenosynovitis  
  crepitans)  
  overview of, 995
- Trihalomethanes  
  use of, 671  
  water contamination with,  
    764–765
- Trinitrotoluene (TNT)  
  liver affected by, 120
- Truman, President, 29
- Tuberculosis  
  deaths, drop in, 485–486  
  screen testing for, 328  
  silicosis accompanied by  
    reactivation of, 1047
- Tumor  
  age of, 161–162  
  detection of, 162  
  doubling times for, 161–162
- Tumor suppressor genes  
  inhibition of, 159
- Tumor suppressor mechanisms  
  DNA-damaged cell replication  
    blocked by, 160<sup>f</sup>
- 20th Century disease  
  term, 938
- UDP-glucuronic acid  
  glucoronid transferase action on,  
    88–89
- Ulcers  
  toxins as cause of, 122
- Ulnar nerve entrapment  
  overview of, 1001
- Ultraviolet radiation  
  overview of, 532–534
- UN Convention on the Rights of the  
  Child, 1556
- UN Habitat  
  environmental health involvement  
    of, 1577
- Uncertainty factor (UF)  
  defined, 444
- Underground mining  
  as high-risk occupation, 960
- “Undue hardship” (defined), 1187
- Unemployment insurance  
  workers’ compensation *versus*, 1340,  
    1341
- Unfit  
  defined, 1141
- “Unfunded liabilities”  
  definition and overview of, 1340
- UNICEF (United Nations Children’s  
  Fund)  
  child labor trends documented by,  
    1556  
  children’s health issues, involvement  
    in, 806  
  environmental health involvement  
    of, 1577

- Unionized plants  
occupational health services in,  
1275, 1276
- United Kingdom  
disability law in, 1133  
occupational health movement in, 24  
public health movement in, 23  
workers' compensation in, 27, 1328
- United Nations, 1576, 1578  
disability as defined by, 1174
- United Nations Development  
Programme (UNDP)  
environmental health involvement  
of, 1578
- United Nations Environment  
Programme (UNEP)  
environmental health involvement  
of, 1578
- Universal Declaration of Human  
Rights, 1948, 1576
- Universal design (concept), 1184
- Upper extremity disorders  
ergonomic factors in, 971
- Upper extremity impairments  
rating of, 1181–1182
- Upper respiratory tract  
particle deposition in, 132
- Upper respiratory tract infection  
lower respiratory tract impacted by,  
143  
prevention of, 1127
- Uranium  
kidneys affected by, 117
- Urbanization, 1550–1551
- Urea formaldehyde foam insulation  
(UFFI)  
use and restrictions on, 714
- Urine  
benzene in, measuring, 96  
drug testing, 1190, 1194,  
1195–1196
- U.S.-Canada border, health issues  
along, 1565
- U.S. Centers for Disease Control and  
Prevention  
tracking terminology used by,  
198–199
- U.S. Chemical Safety and Hazard  
Investigation Board (CSB),  
780–781
- U.S.-Mexico border, health issues  
along, 1564, 1565–1569
- U.S.-Mexico Border Health Associa-  
tion, 1569
- U.S.-Mexico Border Health Commis-  
sion, 1477–1478
- U.S. National Committee for Injury  
Prevention and Control  
injury prevention recommendations  
of, 376, 376*t*
- U.S. Preventive Services Task Force  
(USPSTF)  
purpose of, 1212–1213
- U.S. Public Health Service  
occupational health branch of, 38
- U.S. Surgeon General, health promo-  
tion role of, 1221–1222
- Utilities industry, injury rates in, 952
- Venezuela, occupational medicine in,  
1247
- Ventilation  
evaluation of, 729  
inadequate as sick building  
syndrome (SBS) factor, 721  
indoor air pollution exposure, role  
in limiting, 755  
overview of, 718–719  
*See also* Heating, ventilation, and air  
conditioning (HVAC) system
- Vertebral fracture history as low-back  
pain factor, 974, 976–977

- Vibration  
 effects and disorders, 564–568  
 as low-back pain factor, 985  
 as skin disorder cause, 1030
- Vietnam War veterans  
 psychogenic outbreaks and, 727
- Villermé, Louis-René (1782–1863), 18
- Vinyl chloride  
 hemangiosarcoma, association with, 121  
 as hepatic angiosarcoma risk factor, 238
- Viral infections  
 respiratory tract affected by, 143
- Virchow, Rudolph, 1484–1485
- Virtually safe dose (VSD)  
 defined, 444
- Visible spectrum  
 light in, 534–536
- Volume of distribution  
 pharmacology and toxicology application to, 83
- Voluntary referrals  
 to employee assistance programs, 1233, 1234, 1235
- Vomiting  
 as toxin response, 121–122
- Wage loss formula for disability evaluation  
 overview of, 1338
- Wal-Mart  
 Hurricane Katrina, aid following, 1363  
 medical clinics at, 1258
- Walgreen's  
 medical clinics at, 1258
- Waste disposal  
 solid and municipal, 768–772, 781–782
- special wastes, 785  
*See also* Hazardous waste disposal
- Wastes. *See* Solid and municipal waste
- Wastewater management  
 overview of, 760–761
- Water-borne disease  
 sources of, 762, 763
- Water contamination  
 ingestion compared to, 797
- Water quality  
 chemical contamination, 763–768  
 drinking water, 759–760  
 microbial hazards, 761–763  
 overview of, 757–759  
 recreational waters, 763  
 wastewater management, 760–761
- Water-soluble substances  
 absorption of, 73–74 (hydrophilic)  
 excretion of, 99
- Weapons of mass destruction, 1384–1387
- Weight control  
 as health promotion program component, 1231
- Weight of evidence, 1412, 1415, 1416, 1417
- Weight reduction  
 as health promotion program component, 1227
- Welders  
 infertility problems in, 1105
- Welding goggles  
 overview of, 874–875
- Wellness  
 and productivity management, 1238–1240  
 programs, workplace-centered, 1238
- White, Kerr, 1219–1220
- Whitehall study of British civil servants, 814
- Whitman, Christie Todd, 1451

- Whole-body vibration  
  effects and disorders, 565–566  
  exposure, extent of, 564
- Why Are Some People Healthy and Others Not?*, 487
- Wilcock, Ann, 814
- Witnesscraft, 1432–1438
- Women  
  reproductive problems in, 174
- Women in medicine  
  railroad medicine and surgery, 25
- Women in workplace  
  absenteeism among, 1165  
  stress, sources of, 820
- Work evaluation counselors  
  duties of, 1254
- Work history  
  gaps in, 282
- Work modification  
  physician evaluations leading to, 1155–1156, 1167  
  *See also* Fit subject to work modification; Workplace modification
- Work organization  
  overview of, 373–374  
  restructuring, 1185  
  stress and, 821–825
- Work performance  
  health relationship to, 51
- Workplace Protection Factor(WPF), 861
- Work-related illness  
  prevention of, 889  
  *See also* Occupational diseases and disorders
- Work-related infectious disease. *See* Occupational infections
- Workaholics  
  burnout risk among, 834–835
- Worker education  
  supporting, 284
- Worker-job mismatch  
  psychological effects of, 820
- Worker privacy  
  risks to, 328–329
- Worker recruitment or selection bias  
  as prospective (cohort) study bias, 235–236
- Workers  
  education level and right to know, 422, 424–426  
  risks, communication to, 422, 424–429  
  safety, attitudes concerning, 363–364, 372–373  
  *See also* Injured workers
- Workers, treating  
  in railroad industry, 26
- Workers' compensation  
  abuses in, 1335–1336  
  accountability in, 53  
  acute and primary care role in, 1279
- Americans with Disabilities Act  
  coverage of, 1133  
  appeals, 1334–1337  
  apportionment, 1341–1345  
  attitudes concerning, 926, 936, 1506  
  boards, procedures of, 928–929  
  cancer-related claims, 1092  
  claims and claims management, 955, 1331–1332  
  design and principles of, 1317–1323  
  determining, basis for, 1019  
  disability, calculating, 947  
  documentation in, 1250  
  downsizing impact on, 37  
  enactment of, 27  
  federal regulatory policy in, 949  
  fees for, 1268  
  health record confidentiality, risks to, 329

- history of, 1316, 1323–1331  
impairment assessment in,  
    1177–1179  
impairment assessment outside,  
    1177–1179  
impairment assessment role in,  
    1171–1172, 1176  
insurance, 1339–1341  
medical insurance *versus*, 40  
for occupational disease, 1345–1349  
overview and scope of, 1315–1317  
practitioners serving system of,  
    1246  
records subpoenaed in cases of,  
    1285  
reform of, 52, 1329–1331,  
    1349–1354  
reports and claims, 277–278  
subjective and objective findings,  
    929–931  
term definitions in, 949  
tertiary prevention relationship to,  
    471  
tort litigation in, 1394–1395  
understanding, importance of, 1251
- Workers with disabilities  
    accommodation for, 237–238  
    assistants, personal for, 1186  
    increase in, 287  
    workplace design for, 1184
- Workforce-general population  
    relationship  
    schema of, 341f
- Working class, treating  
    U.S. *versus* European attitudes  
        toward, 22
- Working-class groups  
    health promotion program  
        influence limited for, 1228
- Working hours  
    as burnout factor, 834
- Working population  
    changes in, 287
- Workplace  
    access barriers, removing,  
        1185–1186  
    carcinogens in, 1091*t*–1092*t*  
    definition and overview of, 4  
    disability prevention in, 471  
    exposure in, 70  
    infections, emerging in, 39  
    intervention in, 915–923  
    low *versus* high-risk, 4  
    psychosocial issues in, 692  
    threats, response to, 1357
- Workplace exposure  
    cigarette smoking combined with,  
        227–228
- Workplace hazards  
    studies of, 232, 233  
    *See also* Occupational hazards
- Workplace modification  
    examples of, 1185  
    organizations specializing in,  
        1187–1188
- Workplace risk assessment  
    defined, 876  
    and management, 880–885,  
        880–885
- Worksite health promotion programs  
    components of, 1228–1232, 1227*t*  
    employee assistance programs com-  
        pared to, 1236  
    movement in, 1206  
    overview of, 1222–1223, 1225–1228
- World Health Organization (WHO)  
    children's health issues, involvement  
        in, 806  
    disability as defined by, 1174  
    drinking water led standard of, 766  
    environmental health involvement  
        of, 1576–1578

- World Health Organization (WHO)  
*(continued)*  
formaldehyde limits set by, 715  
handicap as defined by, 1175–1176  
health promotion role of, 1221,  
    1222  
influenza pandemics, potential  
    monitored by, 1381  
travel medicine information  
    provided by, 1571
- World peace  
    physicians as advocates for, 45
- World recession  
    impact of, 1542
- World Trade Center, attack on,  
    1359–1360, 1451, 1458
- World Trade Organization  
    environmental health involvement  
        of, 1577
- World War I  
    induction examinations in, 328
- World War II  
    induction examinations in, 328  
    occupational medicine during, 12,  
        27, 29  
    population health concerns during,  
        1205  
    safety theory in, 367  
    workers' compensation during, 1328
- Wright, Frank Lloyd  
    houses built by, 511  
    overview of, 506–514
- Written communication  
    literacy levels in, 422
- Xenobiotics  
    absorption of (*see* Absorption of  
        xenobiotics)  
    behavior and disposition of, 66,  
        68–70, 69*f* (*see also* Absorption of  
        xenobiotics; Distribution of  
        xenobiotics; Excretion of xenobi-  
            otics; Metabolism of xenobiotics)  
    biotransformation pattern for, 84*f*  
    bloodstream entry rate of, 72–74  
    circulating blood levels, cumulative  
        of, 79*f*  
    definition and overview of, 56,  
        65–66  
    experimental study of, 70  
    ionization of, 74  
    kinetics of, 99–105  
    metabolism of, 72  
    plasma and tissue levels, post-  
        absorption of, 78, 78*f*  
    plasma concentration curve for, 66,  
        68*f*  
    toxicity of, 72
- Xylenes  
    overview and health effects of, 680
- Y2K scare, 1383
- Yellow fever, 1571, 1578  
    occupational exposure and preven-  
        tion, 1126
- Young workers  
    safety, attitudes concerning,  
        372–373
- Zambia, economic and health condi-  
    tions in, 1538–1539
- Zeolite  
    minerals resembling, 578
- Zero-order kinetics  
    conditions resulting in, 102  
    implications of, 94  
    overview of, 93  
    in saturated capacity, 103
- Zoonoses, 1126  
    as occupational hazard, 1124

# **ABOUT THE AUTHOR**

TEE L. GUIDOTTI, MD, MPH is currently an international consultant in occupational and environmental health and medicine. He is the former Professor and Chair of the Department of Environmental and Occupational Health in the School of Public Health and Health Services, The George Washington University Medical Center, Washington D.C., and Director of the Division of Occupational Medicine and Toxicology in the Department of Medicine, School of Medicine and Health Sciences. Prior to taking that position in 1999, he was for fourteen years Professor of Occupational and Environmental Medicine and Director of the Occupational Health Program in the Department of Public Health Sciences at the University of Alberta in Edmonton, Canada. He completed medical school at the University of California at San Diego, trained at the Johns Hopkins Hospital, and obtained his degree in public health at the Johns Hopkins School of Hygiene and Public Health. He holds board certification in the United States in internal medicine, pulmonary medicine, and occupational and environmental medicine and a fellowship in occupational medicine in Canada and the United Kingdom. He also holds professional credentials in toxicology (DABT) and environmental science (QEP). His research interests include occupational and environmental lung diseases and inhalation toxicology. His career has been unusually balanced in occupational and environmental medicine. He is best known in occupational medicine for his expertise on the occupational health problems of firefighters and oil and gas workers. His environmental interests include air quality, risk science, ecosystem and human health, and child

## **About the Author**

health and the environment. He has published over 250 papers and book chapters and produced six books. He has written extensively on the evaluation of scientific evidence in litigation, adjudication, and policy. Dr. Guidotti has served as President of the American College of Occupational and Environmental Medicine and holds many fellowships and awards. He is Editor-in-Chief of the *Archives of Environmental and Occupational Health*. He is the recipient of many awards and honors.