LILIENFELD'S

Foundations of Epidemiology

FOURTH EDITION

Dona Schneider
David E. Lilienfeld

OXFORD

Lilienfeld's Foundations of Epidemiology

FOURTH EDITION

Lilienfeld's Foundations of Epidemiology

DONA SCHNEIDER, PH.D., M.P.H., F.A.C.E.

DAVID E. LILIENFELD, M.D., M.S. ENGIN., M.P.H., F.A.C.E.

Original Edition by **ABRAHAM M. LILIENFELD, M.D., M.P.H., D.SC. (HON.), F.A.C.E.**

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of
Oxford. It furthers the University's objective of excellence in research,
scholarship, and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
Foundations of epidemiology.
Lilienfeld's foundations of epidemiology / Dona Schneider, David E. Lilienfeld; original
edition by Abraham M. Lilienfeld. — Fourth edition.
p. ; cm.
Preceded by Foundations of epidemiology / revised by David E. Lilienfeld,
Paul D. Stolley. 3rd ed. 1994.
Includes bibliographical references and index.
ISBN 978–0–19–537767–5
I. Schneider, Dona, 1946–. II. Lilienfeld, David E. 1957–. III. Title.
[DNLM: 1. Epidemiologic Methods. 2. Epidemiology. WA 950]
RA651
614.4—dc23
2015006834

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

Abraham Morris Lilienfeld (1920–1984), University Distinguished Service Professor at the Johns Hopkins University School of Hygiene and Public Health, wrote the first edition of this book in 1975/6. That edition reflected Lilienfeld's lifelong focus on teaching. One former student described him as having been born "with a piece of chalk in his hand," and with that chalk he would paint rainbows of epidemiology across the blackboard.

Students always had priority in scheduling meetings with Lilienfeld. When those meetings came at the end of the day, he would, at the last minute, regularly invite students home to continue the discussion over dinner. A blackboard hung in the family's kitchen to the side of the informal dining table for precisely this purpose. Doctoral students also sat at that table to review their thesis work with him, or, if more space was required, at the bigger table in the dining room.

Lilienfeld went to great lengths to teach and advocate for epidemiology in any and all venues. He was one of the founders of the University of Minnesota at Minneapolis Summer Program, where he taught epidemiology for three weeks each year. When Alexander Langmuir, his mentor, left Johns Hopkins to found the Epidemic Intelligence Service (EIS) at the Centers for Disease Control, he called upon Lilienfeld to teach introductory epidemiology to all EIS officers during the program's early years. When asked at the last minute to entertain a prominent Kenyan government economics official for the university president, Lilienfeld brought the guest home to "meet the family." During the evening, he regaled the visitor with tales about John Snow and the cholera outbreak investigations. He brought out and showed his guest a piece of the wooden water pipe involved in the outbreak that he husbanded that year as President of the American Epidemiological Society. A few years later, when the Kenyan official's son came to the United States after receiving his MD degree, he attended Johns Hopkins to study epidemiology.

Lilienfeld organized the Public Health Option in the pre-baccalaureate program at Johns Hopkins during the early 1970s, volunteering to teach an introductory epidemiology class for seniors. When the *American Journal of*

Epidemiology was struggling financially in the late 1960s, Lilienfeld suggested that the Society for Epidemiological Research (SER) might be interested in sponsoring the journal. The society subsequently did so. Lilienfeld took great pride in how many SER members learned about different aspects of epidemiology in the course of reading the journal. And, when the first edition of this text appeared, this consummate teacher insisted on keeping the price of the book as low as possible to facilitate access by all students, from undergraduates to post-doctoral fellows. The book appeared at a price of $19.95.

Abraham M. Lilienfeld embodied devotion to teaching about epidemiology to all who would listen. It is to his memory that we dedicate this book, in the spirit of introductory epidemiology as the cornerstone of public health and medicine.

CONTENTS

FOREWORD

The first edition of this text appeared in 1976, but, as explained to me on multiple occasions by the author (my father), its origins trace to more than two decades before. Abraham Morris Lilienfeld trained as an epidemiologist with a focus on infectious diseases. On completion of his training, he joined the epidemiology faculty at the Johns Hopkins School of Hygiene and Public Health (JHU), where he became interested in liver cirrhosis following his work on an infectious hepatitis outbreak in the early 1950s. The etiology of liver cirrhosis was not well understood at that time, but the prevalent hypothesis attributed the disease to alcoholism—not to an infectious agent. Lilienfeld examined the alcohol–liver cirrhosis relationship using the only epidemiologic methods he had available—those traditionally used to examine the etiology of infectious diseases. The utility of those methods for studying noninfectious diseases quickly became evident to him.

When Lilienfeld assumed leadership of the epidemiology program at Roswell Park Memorial Cancer Institute in Buffalo, New York, in 1953, he brought with him a plan to apply traditional epidemiologic methods to study the etiology of cancer. He conceptualized the issue of breast cancer mortality in a new way, with susceptibility to the disease varying with menopausal status, similar to the way susceptibility to an infectious agent varies based on immunization status. This change in framing the question allowed him to demonstrate the benefit of using traditional epidemiologic tools with the emerging diseases of the time—heart disease, cancer, and stroke.

Lilienfeld returned to JHU with a reputation as a methodological innovator. In 1965 the World Health Organization Cancer Unit asked him to collaborate on a book for health care professionals on epidemiologic methods for the study of cancer. *Cancer Epidemiology: Methods of Study* (1967) includes a discussion of John Snow and the mid-Victorian cholera outbreaks in the first five pages of the first chapter. Lilienfeld used this book in his courses on chronic disease epidemiology at JHU for several years. He noted that students found it helpful to have the material presented as a story, and he used the development of epidemiologic knowledge about cigarette smoking and lung cancer as his narrative.

In 1970, Lilienfeld became chair of epidemiology at JHU. With that appointment he took over responsibility for the department's introductory course. He taught the concepts and methods of epidemiology as foundational, truly believing the difference between chronic and infectious disease epidemiology to be an artificial one. By 1972, an entrepreneurial editor (Jeffrey House) at the Oxford University Press approached Lilienfeld about using his lecture notes as the basis for an introductory textbook in epidemiology. The professor declined, as he was lobbying Congress for the Surveillance, Epidemiology, and End Results (SEER) system as part of President Nixon's "War on Cancer," helping multiple doctoral students with their thesis research, developing a center for the epidemiology of gastrointestinal diseases, and continuing to teach his courses. He simply didn't have the time needed to write the book.

Two years later, in November 1974, Lilienfeld collapsed with a cardiac arrest during a lecture in the introductory course in epidemiology. After being resuscitated by the students, he found himself in a rehabilitation program and had to be out of the office for months. Needing a project to engage his considerable intellectual energy, Lilienfeld remembered the invitation to write an introductory textbook on epidemiology. An inquiry found the invitation still open. Within six months, the first draft was completed. Thus was born the **first edition** of *Foundations of Epidemiology* (1976). Students responded positively to the book, and it was adopted for many introductory epidemiology courses.

The **second edition** of the text appeared in 1980, a collaborative effort by father and son. The book was updated to reflect students' stated desires for a different narrative, as the smoking and lung cancer saga was becoming dated. As a result, the second edition shifted to narratives about oral contraceptives and cardiovascular diseases. The discussion of randomized controlled trials was also considerably expanded, reflecting epidemiologists' increasing use of them. The book was thoroughly updated, and the cover changed to show the point-contact spread of an infectious disease outbreak (although in production, the figure was mistakenly inverted).

The Lilienfelds were working on a **third edition** when the elder author died in 1984. Work on the revision was shelved until the early 1990s when a collaboration was struck with Paul Stolley. The didactic material in the third edition was broadened to accord with the expansion in epidemiologic activities during the 1980s. Demographic studies were discussed in a defined section, and the same was true of epidemiologic studies. A discussion of epidemiology in clinical practice was added.

For the **fourth edition**, the role of epidemiology within public health is the central theme. The chapter on inferences is moved forward, as many instructors now discuss causality early in their introductory epidemiology courses. The examples are updated, discussions about vital statistics are broadened to include birth as well as death, and recent concepts such as quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs) have been added. To keep the book brief, it is purposely lean on figures and diagrams. To keep it affordable for students, it has no photos or color sidebars.

Those who have read previous editions of this text will find an old friend in this new edition, one you've known well for a long time, perhaps with some new attire and new stories to be shared. Hopefully, it succeeds in remaining true to the vision Lilienfeld had when creating the first edition describing an integrated field known as epidemiology. While today's world is different from that which he inhabited, and while the field itself has evolved, the core of the information remains as *Lilienfeld's Foundations of Epidemiology*.

DEL

PREFACE

This book is designed as a foundational text for introductory courses in epidemiology wherever they are offered—in schools of public health, medicine, dentistry, nursing, and the allied health professions, as well as undergraduate programs offered by two- and four-year liberal arts colleges and technical schools. The original 1976 edition of this text was written for precisely that purpose, and the fourth edition continues in the same tradition. We recognize, however, that pedagogical expectations for an introductory epidemiology text have changed over the years. Today, at the undergraduate level a successful text must address the set of expected discipline-specific learning outcomes developed from a national consensus of public health and liberal arts educators. At the graduate level such a text must help students develop the epidemiologic competencies that will aid them in their professional careers. We believe a foundational text can do both.

Because epidemiology is an inherently integrative discipline, it is not uncommon for texts to use different words to describe the same concept. For instance, some texts use the term *relative risk*, whereas others may use *rate ratio* or *risk ratio*. It can be confusing for students if their reading of the literature is not in line with the teachings in their introductory text. To address this dissonance, throughout this text complementary terms appear in parentheses and all definitions provided are synonymous with those from the *Dictionary of Epidemiology, 6th Edition*.

This text is divided into four parts. Part I (Introduction to Epidemiology) reviews the historical background and conceptual basis for epidemiology. Part II (Descriptive Studies) covers sources of epidemiologic data and the study designs that can be used to describe mortality and morbidity in human populations. The epidemiologic designs used to test hypotheses about health-related outcomes are discussed in detail in Part III (Analytic Studies), including the advantages and disadvantages of each. Finally, Part IV (Using Epidemiologic Information) includes chapters that show the reach of epidemiology into various disciplines—including field investigations, health care planning, and the clinical realms.

The reasoning processes used by epidemiologists to address health-related problems are illustrated throughout the book using both classic and contemporary examples. Problems sets are provided in some chapters to give students an opportunity to apply the epidemiologic methods and reasoning processes that constitute the field to contemporary health-related issues. Some of these issues are intentionally designed to evoke various viewpoints and should serve to spark classroom discussion. Thus, the text covers a broad reach of epidemiologic concepts and focuses on the interdisciplinary approach. We feel it important to describe not only how epidemiologic concepts developed over time, but also to demonstrate how the approach continues to be practical for understanding today's, and even tomorrow's, health-related outcomes.

I | Introduction to Epidemiology

The work of epidemiology is related to unanswered questions, but also to unquestioned answers.[1]

Patricia Buffler (2011)

PART I PROVIDES AN introduction to epidemiologic thinking and how it developed over time. The opening chapter explains how epidemiology uses a *comparative* approach, focusing on *disparities* (how health-related outcomes vary across time, in different places, and among different population subgroups). The epidemiologist's ultimate goal is to find, reduce, or eliminate those factors that cause disease and other adverse health outcomes. To achieve this goal, the epidemiologist must first determine who is at risk and why. The second step is for the epidemiologist to conduct carefully designed studies to test hypotheses about potential *etiological* (*causal*) *factors*.

Chapter 2 gives a brief account of the history of epidemiology, showing how epidemiology evolved into its present form. The major figures in the development of epidemiology are mentioned, and historical examples that demonstrate the uses of epidemiology are presented.

Chapter 3 covers the general principles and terminology used to classify diseases for the purposes of epidemiologic studies. The triad of agent–host–environment is discussed in detail, and the spectrum of disease, both infectious and noninfectious, with its associated terminology is introduced. Chapter 3 concludes with a discussion of herd immunity, the epidemiologic basis for national vaccination policies.

This introduction to epidemiology concludes with Chapter 4, which describes the way in which epidemiologists draw inferences from hypothesis-based studies. Of particular importance is how the epidemiologist determines whether the statistical results from these studies do or do not support a factor being causally-related to a particular health outcome.

Reference

1. Patricia Buffler, "Keynote address for the North American Congress of Epidemiology," Montreal, Canada, July 2011.

1 | Laying the Foundations

> Epidemiology came to mean the study of disease, any disease, as a mass phenomenon . . . The physician's unit of study is a single human being . . . The epidemiologist's unit of study is . . . an aggregate of human beings.[1]
>
> *Major Greenwood (1932)*

Epidemiology is "the study of the occurrence and distribution of health-related events, states, and processes in specified populations, including the study of the determinants influencing such processes, and the application of this knowledge to control relevant health problems."[2] In other words, not only does epidemiology identify patterns of health-related problems in populations, but it also investigates the underlying causes of those problems and offers the results of well-designed studies as the basis for implementing plans to improve the public's health. We must add that epidemiology is an integrative, eclectic science deriving concepts and methods from other disciplines, especially anthropology, biology, geography, history, sociology, and statistics. This interdisciplinary approach has led to epidemiology being taught not only in medical schools and schools of public health, but also to undergraduates in both two-year and four-year liberal arts and professional programs such as nursing and the allied health professions.

Epidemiologists are primarily interested in the way health outcomes differ according to *time, place*, and *persons*. They examine whether changes have occurred in health-related states or events over days, months, or years (time); whether one geographical area (place) differs from another in the frequency of health-related outcomes; and whether the characteristics of individuals (persons) with a particular disease or condition distinguish them from others.

Epidemiologists are concerned with the following characteristics of persons:

- Demographic factors such as age, gender, race, and ethnic group
- Biological factors such as circulating levels of antibodies, chemicals, and enzymes; blood constituents such as cells and platelets; and measurements of the physiological functions of organs and systems, such as blood glucose or hormone levels
- Social and economic factors such as socioeconomic status, educational background, occupation, and place of birth
- Behavioral factors such as tobacco and drug use, diet, and physical exercise
- Genetic factors such as blood groups or gene mutations

Understanding the confluence of these characteristics of persons in time and space, and being able to define each of them clearly and precisely at the outset of an epidemiologic inquiry, is the cornerstone to developing a good study design.

Purposes of Epidemiologic Inquiries

Epidemiologic studies are usually designed to yield information that can:

- Provide data that will help clarify the *etiology* or cause(s) of specific health outcomes
- Determine whether epidemiologic data are consistent with causal hypotheses developed clinically, experimentally, or from other studies
- Provide data that can help develop and evaluate preventive procedures, public health practices, and other types of health-related services

Etiological Studies

An example of how epidemiologists seek to determine *etiologic* (causal) *factors* comes from occupational epidemiology, where the frequency of disease may be observed to be higher among workers with particular exposures than among the general population. For instance, in 1955 the English epidemiologist Sir Richard Doll reported the occurrence of 18 cases of lung cancer among 105 asbestos factory workers. He concluded that asbestos was a cause of lung cancer, although the risk seemed confined to those with the greatest exposures.[3]

Selikoff and his colleagues examined the effects of asbestos on the end users of asbestos products by reviewing two decades of records from building insulation workers' unions in New York and New Jersey. Beginning on December 31, 1942, the deaths of members were followed to see what proportion died from lung cancer. The researchers found that of 255 union members, 45 died from cancers of the lung and pleura. They concurred with Doll that asbestos was a cause of lung cancer, even for workers who had lower exposures to asbestos than those experienced by the factory workers.[4]

Could cigarette smoking explain some of the relationship between asbestos and lung cancer? The investigators addressed this question by expanding their previous study to include all 17,800 asbestos insulation workers in the United States. Using data from a study conducted by the American Cancer Society to determine the rate of death from lung cancer in the general population, the researchers determined that asbestos insulation workers who smoked cigarettes were 53.2 times as likely to die from lung cancer compared with nonsmokers with no exposure to asbestos. For workers exposed to asbestos who did not smoke cigarettes, the elevation in risk was 5.2 times as high. Among the general population (those not exposed to asbestos in the workplace), cigarette smokers had a risk of lung cancer that was 10.8 times that of nonsmokers. The researchers concluded that asbestos was a cause of lung cancer, and that exposure to asbestos interacts with cigarette smoking to markedly increase that risk beyond that of either factor alone.[5]

It is not always true that exposure to chemicals increases the risk of adverse health outcomes. Indeed, occasionally an epidemiologic study finds that a chemical exposure provides a protective effect. A classic example of a protective effect is that of the presence of fluoride in drinking water reducing the risk of dental caries. Around 1915, a practicing dentist in Colorado formed the clinical impression that his patients with mottled teeth had fewer dental caries than his other patients.[6–8] By the late 1930s, dentists understood that patients presenting with mottled tooth enamel had been chronically exposed to high levels of fluoride in drinking water before their permanent teeth had erupted.[9–11] This combination of findings led the Public Health Service to conduct surveys of children aged 12–14 years in 21 cities in four states where the fluoride concentration in the water supply varied considerably.[12] The results showed that dental caries decreased with increasing fluoride content in drinking water, suggesting that adding fluoride to the water supply should decrease the frequency of dental caries (Figure 1.1).

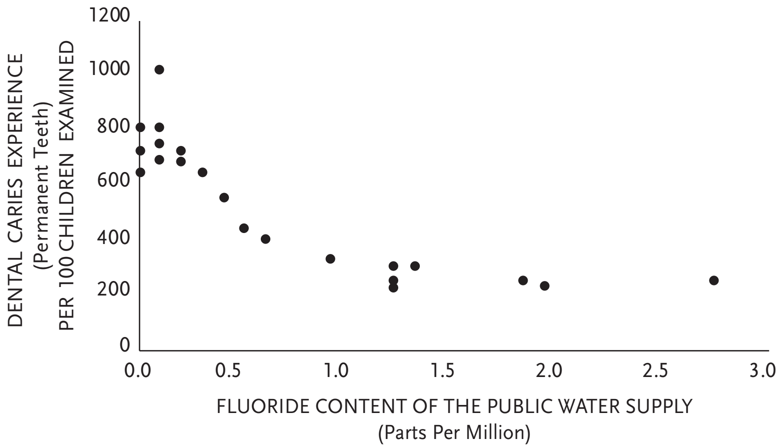

FIGURE 1.1 Relationship between the amount of dental caries (permanent teeth) observed in 7257 selected 12- to 14-year-old white school children of 21 cities of 4 states and the fluoride (F) content of public water supply.
SOURCE: Dean, Arnold, and Elvove (1942).[12]

The relationship between exposure to fluoride and reduced dental caries was then tested by experiment. The Public Health Service determined that fluoride would be added to the water supply of one community, and the water supply of a comparable community would remain naturally low. Over the course of several years, dentists could record the dental caries experience of school children in the communities and the rates could be compared. Several matched community studies were initiated, including one in 1945 comparing Newburgh and Kingston, New York.[13] After several years, school children in Newburgh (the town with fluoride added to the water supply) had 50 percent fewer decayed, missing, or filled teeth than school children in Kingston (the town without added fluoride). This example demonstrates how epidemiologic information is used to develop public health policy; namely, a practitioner's clinical impression led to an epidemiologic survey that then led to a well-designed epidemiologic experiment. The results of the experimental study eventually led to public health policy whereby fluoride was added to the drinking water supplies of many US communities.

Evaluating Consistency

It is fairly common for an epidemiologist to test an etiological hypothesis developed from clinical, experimental, or other studies to see whether it is consistent, i.e. whether the results will be the same if the study is repeated using various populations in other places or at different times. For example, oral contraceptives became widely available during the early 1960s. Case reports and the results of several epidemiologic studies showed relationships between oral contraceptive use and venous thromboembolism (blood clots), thrombotic stroke, and myocardial infarction.[14–16] Additional studies were undertaken to see if the findings held for women in different age groups.[17,18] Scores of epidemiologic studies followed and continue today comparing the risk of contraceptive use among women living in various countries, who have varying risk factors (smoking, family history of stroke), and who are taking different formulations of oral contraceptives, at different dosages, and for different lengths of time. The findings have been consistently demonstrated across multiple study populations over the years.

Basis for Preventive and Public Health Services

Epidemiologic data provide a means of evaluating the current health of populations, whether health-related outcomes vary over time, and whether prevention programs work. Examples of epidemiologic data include birth and death rates, as well as information gathered by disease registries and surveys of health risk behaviors. A complete discussion of these types of data appears in Part II of this text. Disease surveillance programs also add to the wealth of information on population health, as does the evaluation of the effectiveness of public health services. An in-depth discussion of the role of surveillance appears in Chapter 7, and new ways of obtaining surveillance data are covered in Chapter 14.

The study that determined the effectiveness of the Salk vaccine for preventing poliomyelitis is a classic example of how epidemiology contributes to preventive and public health services.[19] A more recent example is how epidemiologic data helped formulate recommendations for the human papilloma virus (HPV) vaccine program. Epidemiologic modeling suggested that vaccinating preadolescent females for HPV would reduce the number of cervical cancer cases in the vaccinated group by 62 percent. Vaccinating boys, however, would only reduce cervical cancer cases by 2 percent, a less cost-effective approach compared to a female-focused vaccination program.[20] Accordingly, HPV vaccination was originally recommended for girls alone. That recommendation was later amended to include boys when new epidemiologic data suggested that vaccinating boys would also reduce the occurrence of 7,000 penile, anal, and oropharyngeal cancers in the United States annually.[21] This example demonstrates how necessary it is to routinely review epidemiologic data in order to best design and implement successful public health prevention programs.

Epidemiologic information on how health status is distributed across time, place, and persons informs public health practice. It allows physicians and public health professionals to target populations where prevention, screening, and healthcare services should be focused in order to get the most out of public health resources. This holds true even if the underlying cause of a particular health state is not known. For example, diabetes mellitus was demonstrated to run in families by researchers at the Mayo Clinic as early as 1952.[22] The reason why this happens, whether there is a genetic etiology for diabetes, or whether environmental factors common to family members explain the development of diabetes, is not really important. All that a physician or public health practitioner needs to know is that there is an increased risk of developing diabetes among the family members of known diabetics. Thus, screening a diabetic patient's parents, siblings, and offspring will likely yield additional cases of the disease in this high-risk population.

Content of Epidemiologic Activities

Epidemiologists engage in two broad areas of study: observational and experimental epidemiology. Each area involves different methods.

Observational Epidemiology

The vast majority of epidemiologic studies fall into the observational category. For instance, the studies described above of employees exposed to certain chemical compounds, surveys of dental caries, and familial aggregation of diabetes, are all examples of observational investigations. Epidemiologists have developed appropriate methods for selecting populations and subgroups to observe, as well as various techniques for analyzing the information obtained from such studies. These topics are covered in Chapters 5 through 8 (Descriptive Studies) and in Chapters 9 and 10 (Observational Studies).

Occasionally, an investigator may observe the occurrence of a disease or other health-related outcome under existing conditions that closely approximate a controlled experiment. Inferences about causal factors derived from these *natural experiments* are considerably stronger than if they had been derived solely from observational studies. The studies in England by Doll, Hill, and Peto on the relationship between tobacco use and lung cancer illustrate this approach.[23–26] In 1951, the investigators ascertained the smoking habits of British male physicians aged 35 years and over and followed them

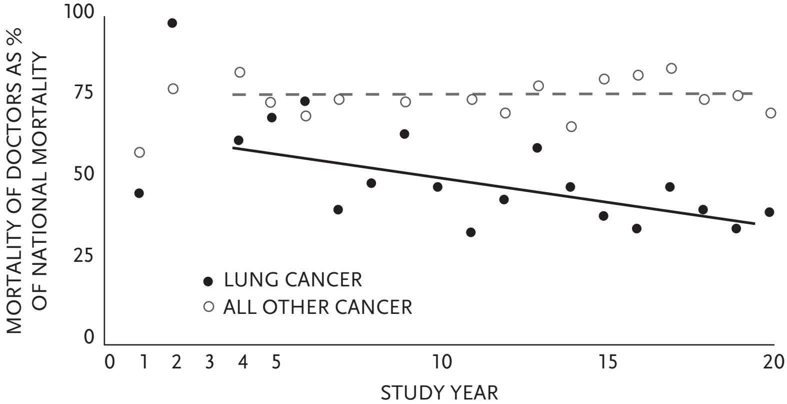

FIGURE 1.2 Trend in number of deaths certified in male doctors as percentage of number expected from experience of all men in England and Wales of same ages.
SOURCE: Doll and Peto (1976).[24] Copyright 1976, *British Medical Journal*. Reproduced with permission of BMJ Publishing Group, Ltd.

to determine their mortality from different causes, particularly lung cancer. Initial findings indicated that physicians who smoked cigarettes had a mortality rate from lung cancer that was about 10 times that of nonsmoking physicians.[23] Questionnaires were sent to the same physicians to determine their cigarette smoking habits in 1956, 1966, and 1971. Data from these surveys showed an approximately 50 percent decline in cigarette smoking among the physicians over the 20-year period.[24] The investigators then compared the cancer mortality experience for the physicians.[23] The mortality from lung cancer declined by about 40 percent with essentially no decline in other cancer deaths. In this case, the decline in cigarette smoking among male British physicians was a natural experiment that yielded a concomitant decline in lung cancer mortality for that group (Figure 1.2).[23]

Experimental Epidemiology

In controlled experiments, the investigator controls which population groups are exposed to specific *therapeutic* interventions (e.g., drugs to treat a disease, surgical procedures, and behavior modification) or *preventive* interventions (vitamin supplementation and smoking cessation programs). The Newburgh–Kingston dental caries study presented above was a controlled experiment. An important feature of experimental studies is that the investigator can randomly allocate subjects to the experimental and control groups, thereby minimizing the influence of factors other than the one being studied in the trial. The methods for assigning subjects in an experimental study are discussed in greater detail in Chapters 11 and 12.

Development and Evaluation of Study Methods

As new public health challenges emerged over time, epidemiologists developed new analytical methods to address them. Some of the methods were adapted from other disciplines, specifically those that utilized a *comparative* approach (comparing different time periods, places, and populations or subgroups). Epidemiologists use the comparative approach for various types of investigations, including those comparing infectious and noninfectious diseases, acute and chronic conditions effecting population health, public health prevention efforts, and medical interventions.[27]

The Sequence of Epidemiologic Reasoning

Epidemiologic reasoning is a two-stage process for clarifying the etiology of health-related states or outcomes. The sequence of reasoning is as follows:

1. Determine the statistical association between a characteristic and a health outcome
2. Derive causal inferences from the patterns of the statistical associations

The methods used to determine the statistical associations may be based on either group or individual factors. While there is a certain degree of overlap between these two categories, it is extremely useful to make the distinction between them. For example, *descriptive studies* (sometimes called *demographic studies*) allow the epidemiologist to compare the health status of different population groups in the expectation that any observed differences can be related to differences in the local environments, personal living habits, or even the genetic composition of these groups. Descriptive studies also provide information on trends in population health and may lead to hypotheses about why the observed patterns exist. The drawback of descriptive studies is that while data on the population being studied may be available, information on the individuals that make up that population may not. In those instances, the epidemiologist may rely on *ecological data* (also called *aggregate data, group data,* or *population data*) for making group comparisons. This lack of information on individuals presents a problem.

As an example, let us assume that Community A (with a high consumption of alcohol) has a higher mortality rate from lung cancer than

Community B (with no consumption of alcohol). This comparison suggests that drinking alcohol may be of etiological importance for developing lung cancer. However, the statistical relationship with alcohol only provides a clue for further investigation because of the problem of *ecological fallacy*, an erroneous conclusion that a statistical relationship existing at the group level also holds at the individual level.[28] While Communities A and B differ in their alcohol consumption, they may also differ in other factors not examined, such as cigarette smoking. One or more of these unexamined factors may be an underlying explanation for the observed lung cancer mortality experiences of the two communities, and not necessarily their alcohol consumption.[29–32]

After an association has been established from ecological data, clinical observations, or laboratory experiments, the epidemiologist attempts to determine whether the same association is present among individuals. Our ecological analysis suggested an association between alcohol consumption and developing lung cancer. Questions posed about individuals from both Communities A and B might include the following:

1. Do persons with the disease (lung cancer) have the characteristic (alcohol consumption) more frequently than those without the disease?
2. Do persons with the characteristic (alcohol consumption) develop the disease (lung cancer) more frequently than those who do not have the characteristic?

These questions can be addressed through cross-sectional, case-control, and cohort study designs, discussed in detail in Chapters 7, 9, and 10.

A Case Study Exemplifying the Epidemiologic Approach

One of the most common uses of the epidemiologic approach is the investigation of a foodborne outbreak. The term *outbreak* is often used to describe an excess of cases in a localized or time-limited situation, whereas the term *epidemic* is often preferred for describing a situation that is more widespread or that occurs over a longer period of time. While the primary goal of foodborne outbreak investigations is to identify the microbe or chemical contaminating the food and causing the illness, the secondary goal is to put safety measures into place (such as removing contaminated foodstuffs from stores, training food handlers to wash their hands, or repairing refrigeration units) to prevent a recurrence.

Foodborne outbreak investigations ascertain (1) whether there has been an outbreak and (2) whether there is a statistical association between the consumption of a specific food and a specific foodborne illness. Consider the classic example of an outbreak of gastroenteritis originally investigated by the New York State Department of Public Health. On an April day in 1940, the local health officer for Lycoming in Oswego County reported an apparent outbreak of acute gastrointestinal illness to the regional health office. An epidemiologist was assigned to investigate and learned that all persons known to be ill had attended a church supper the previous evening. As their family members who had not attended the event had not become ill, the epidemiologist focused on events related to the church supper. All attendees at the supper were interviewed to determine whether or not they developed symptoms of gastroenteritis. If so, they were asked when (day and hour) the symptoms first appeared. Everyone was asked which food(s) they consumed at the supper. Of the 80 attendees, 46 persons were found to have had symptoms of gastroenteritis.[33]

The first step in assessing whether an outbreak has occurred requires calculating the *crude attack rate*.

$$\textit{Crude attack rate} = \frac{\textit{Number of persons ill with disease}}{\textit{Number of persons attending the event}}$$

In this case, the crude attack rate was almost 58 percent (46/80), far beyond what might be expected in the general population (the *endemic* or background rate of disease) or what might occur from seasonal or random variation. This finding led the epidemiologist to conclude that he was dealing with a probable foodborne outbreak.

The second step in an outbreak investigation is to develop a *case definition* that includes time, place, and person variables, as well as diagnosis or symptomatology. In our example, the case definition was as follows: All persons who developed acute gastrointestinal *symptoms* within 72 hours after eating supper on April 18, 1940 (*time*), and who were among attendees (*persons*) of the Oswego Church supper in Lycoming, New York (*place*). Every individual who attended the event was interviewed (whether they became ill or not) to determine what foods (including water and condiments) they did or did not consume. If they became ill, it was important to determine the time when their symptoms first appeared. The information was then entered into a *line listing* (spreadsheet) and used to develop two epidemiologic tools: a table of food-specific attack rates and an *epidemic curve*.

The attack rates for each food served at the church supper were calculated using the following formula:

$$Food\text{-}specific\ attack\ rate = \frac{Number\ of\ persons\ who\ ate\ the\ food\ and\ became\ ill}{Total\ number\ of\ persons\ who\ ate\ the\ food}$$

To identify the contaminated food, the epidemiologist then calculated the ratio of food-specific attack rates (*rate ratio*) for each menu item and compared the results to identify which food(s) were the likely cause of the illness.

$$Rate\ ratio = \frac{Food\text{-}specific\ attack\ rate\ for\ those\ who\ become\ ill}{Food\text{-}specific\ attack\ rate\ for\ those\ who\ remained\ well}$$

Table 1.1 shows that persons who ate the vanilla ice cream at the church supper were 5.7 times more likely to become ill than those who did not eat it, representing the greatest discrepancy between the attack rates of those consuming and not consuming each specific foodstuff. The epidemiologist was able to conclude that the vanilla ice cream was the likely source of the contamination that caused the illness.

Laboratory investigations may be undertaken to identify the chemical or microorganism that contaminated the food and was responsible for the outbreak. Samples from remaining food items may further inform the investigation, as well as biological samples (stool and/or blood) from individuals who became ill and from those who prepared or served the food. It may be impossible to sample the food identified as the source of the outbreak if it was completely consumed or discarded. In these instances, the investigator may make efforts to track food items back into the food supply or use an additional tool—the epidemic curve—to aid in the identification of the most likely agent. A full discussion of the epidemic curve is presented in Chapter 3.

Foodborne and other types of outbreaks may be difficult to investigate because the world has become a complex place. Interstate and international travel, trade agreements, and the rapid transport of goods and foodstuffs can make it difficult to work through foodborne outbreaks. When the outbreak described above occurred in 1940, surveillance systems with real-time reporting did not exist. In 1996, the United States Centers for Disease Control and Prevention (CDC) put into place *FoodNet*, an active surveillance network covering 10 areas of the United States (Connecticut, Georgia, Maryland, Minnesota, New Mexico, Oregon, Tennessee, and selected counties in California, Colorado, and New York).[34] FoodNet requires reporting

TABLE 1.1 Food-Specific Attack Rates

FOOD	ATE THE FOOD				DID NOT EAT THE FOOD				RATE RATIO*
	ILL	WELL	TOTAL	ATTACK RATE (%)	ILL	WELL	TOTAL	ATTACK RATE (%)	
Baked ham	29	17	46	63	17	12	29	59	1.1
Spinach	26	17	43	60	20	12	32	63	1.0
Mashed potato	23	14	37	62	23	14	37	62	1.0
Cabbage salad	18	10	28	64	28	19	47	60	1.1
Jello	16	7	23	70	30	22	52	58	1.2
Rolls	21	16	37	57	25	13	38	66	0.9
Brown bread	18	9	27	67	28	20	48	58	1.2
Milk	2	2	4	50	44	27	71	62	0.8
Coffee	19	12	31	61	27	17	44	61	1.0
Water	13	11	24	54	33	18	51	65	0.8
Cakes	27	13	40	68	19	16	35	54	1.3
Ice cream (van)	43	11	54	80	3	18	21	14	5.7
Ice cream (choc)	25	22	47	53	20	7	27	74	0.7
Fruit salad	4	2	6	67	42	27	69	61	1.1

* RATE RATIO IS CALCULATED AS RATIO OF CALCULATED ATTACK RATES. FOR INSTANCE, FOR BAKED HAM, THE RATE RATIO IS 0.63/0.59, OR 1.1.

SOURCE: Centers for Disease Control and Prevention (1981).[33]

of all laboratory-confirmed cases of infections caused by *Campylobacter, Cryptosporidium, Cyclospora, Listeria, Salmonella,* Shiga toxin-producing *Escherichia coli* (STEC 0157 and non-0157), *Shigella, Vibrio*, and *Yersinia*, as well as hospitalized cases of hemolytic uremic syndrome (a complication of STEC). FoodNet does not, however, serve the purpose of national outbreak surveillance in real time. To that end, the CDC maintains the *National Outbreak Reporting System* (NORS)[35] to track all foodborne, waterborne, and other types of enteric outbreaks in all US states and territories. NORS became Web-based as of 2009, allowing faster responses to outbreaks by state and local public health agencies. The data generated from NORS should allow epidemiologists and health policymakers to design and implement more effective measures to reduce the burden of foodborne and waterborne outbreaks on the population. Similar national foodborne disease surveillance systems have been developed in the United Kingdom (by the Health Protection Agency), France (a network of 14 National Reference Centers), and other European countries, under the aegis of the World Health Organization.

The interested reader can find information on the historical use of the Lycoming outbreak in Gross's description of the outbreak investigation 35 years after it occurred.[36] The techniques used in the Lycoming outbreak investigation have now been used by epidemiologists for more than a century, although the statistical analyses used have evolved since the first known effort to investigate a foodborne outbreak in England in 1902.[37]

Summary

Epidemiology is a comparative science in which the investigator examines the relationship of health-related outcomes with the presence or absence of various factors in populations. Descriptive studies are used to formulate etiological hypotheses about the patterns of these outcomes based on time, place, and person factors. Analytic studies (observational and experimental) are used to test hypotheses and to provide insights into the potential causes of the outcomes being evaluated. Epidemiologic activities include (1) descriptive and observational studies, where the investigator does not control exposure to subjects in the study, (2) experimental studies, where the investigator does control subjects' exposure, as well as (3) the development and evaluation of new study methods.

Epidemiologic reasoning is a two-stage process for clarifying the etiology of health-related states or outcomes. The first stage is deriving a statistical association between a characteristic and a health outcome, a task that may be achieved through the use of descriptive studies. Conclusions based on statistical relationships derived from descriptive studies must be formulated carefully, however, as they are based on population data and are prone to ecological fallacy.

The second stage of epidemiologic reasoning requires testing causal hypotheses (generated from the first stage) using analytic methods. An example of the second stage was offered in the investigation of a foodborne outbreak. The source of an outbreak could be found by comparing the attack rates among persons who consumed a given food with those for persons who did not consume that food. The food item with the greatest disparity in attack rates (rate ratio) was the likely source of the outbreak.

References

1. Major Greenwood, *Epidemiology: Historical and Experimental*. Baltimore: Johns Hopkins Press, 1932.
2. Miquel Porta, ed., *A Dictionary of Epidemiology, 6th Edition*. New York: Oxford University Press, 2014.
3. Richard Doll, "Mortality from Lung Cancer in Asbestos Workers," *British Journal of Industrial Medicine* 12 (1955): 81–86.
4. Irving J. Selikoff, Jacob Churg, and E. Cuyler Hammond, "Asbestos Exposure and Neoplasia." *JAMA* 188 (1964): 22–26.
5. Irving J. Selikoff, E. Cuyler Hammond, and Herbert Seidman, "Mortality Experience of Insulation Workers in the United States and Canada, 1943–1976," *Annals of the New York Academy of Medicine* 330 (1979): 91–116.
6. F. S. McKay and G. V. Black, "Mottled Teeth: An Endemic Developmental Imperfection of the Enamel of the Teeth Heretofore Unknown in the Literature of Dentistry," *Dental Cosmos* 58, no. 5 (1916): 477–484.
7. F. S. McKay and G. V. Black, "An Investigation of Mottled Teeth," *Dental Cosmos* 58 (1916): 627–644; 781–792; 894–904.
8. F. S. McKay, "Mottled Enamel: A Fundamental Problem in Dentistry," *Dental Cosmos* 67, no. 9 (1925): 847–860.
9. H. Trendley Dean and Elias Elvove, "Some Epidemiological Aspects of Chronic Endemic Dental Fluorosis," *American Journal of Public Health and the Nation's Health* 26, no. 6 (1936): 567–575.
10. H. Trendley Dean, "Endemic Fluorosis and Its Relation to Dental Caries," *Public Health Reports* 53, no. 33 (1938): 1443–1498.
11. H. Trendley Dean, Phillip Jay, Francis A. Arnold, Jr., Frank J. McClure, and Elias Elvove, "Domestic Water and Dental Caries, Including Certain Epidemiological Aspects of Oral *L. acidophilus*," *Public Health Reports* 54, no. 21 (1939): 862–888.
12. H. Trendley Dean, Francis A. Arnold, and Elias Elvove, "Domestic Water and Dental Caries," *Public Health Reports* 57, no. 32 (1942): 1155–1179.
13. David B. Ast and Edward R. Schlesinger, "The Conclusion of a Ten-Year Study of Water Fluoridation," *American Journal of Public Health and the Nation's Health* 46, no 3 (1956): 265–271.

14. J. Boyce, J. W. Fawcett, and E. W. Noall, "Coronary Thrombosis and Conovide," *Lancet* 1, no. 7272 (1963): 111.
15. Collaborative Group for the Study of Stroke in Young Women. "Oral Contraception and Increased Risk of Cerebral Ischemia or Thrombosis," *New England Journal of Medicine* 288, no. 17 (1973): 871–878.
16. M. P. Vessey and J. I. Mann, "Female Sex Hormones and Thrombosis. Epidemiological Aspects," *British Medical Bulletin* 2, no. 34 (1978): 157–162.
17. J. I. Mann, Richard Doll, Margaret Thorogood, M. P. Vessey, and W. E. Waters, "Risk Factors for Myocardial Infarction in Young Women," *British Journal of Preventive & Social Medicine* 30, no. 2 (1976): 94–100.
18. J. I. Mann, M. H. W. Inman, and Margaret Thorogood, "Oral Contraceptive Use in Older Women and Fatal Myocardial Infarction," *British Medical Journal* 2, no 6033 (1976): 445–447.
19. Francis T Jr, Korns RF, Voight RB, et al. An Evaluation of the 1954 Poliomyelitis Vaccine Trials. *American Journal of Public Health and the Nation's Health* 45, Part 2 (1955): 1–63.
20. A. V. Taira, C. P. Neukermans, and G. D. Sanders, "Evaluating Human Papillomavirus Vaccination Programs," *Emerging Infectious Diseases* 10, no. 11 (2004): 1915–1923.
21. Centers for Disease Control and Prevention. "Vaccines: HPV Vaccine FAQS." Available at http://www.cdc.gov/vaccines/vpd-vac/hpv/vac-faqs.htm.
22. Arthur G. Steinberg and Russell M. Wilder, "A Study of the Genetics of Diabetes Mellitus," *American Journal of Human Genetics* 4, no. 2 (1952): 113–135.
23. Richard Doll and A. Bradford Hill, "Smoking and Carcinoma of the Lung," *British Medical Journal* 2, no. 4682 (1950): 739–748.
24. Richard Doll and Richard Peto, "Mortality in Relation to Smoking: 20 Years' Observations on Male British Doctors," *British Medical Journal* 2, no. 6051 (1976): 1525–1536.
25. Richard Doll, Richard Gray, Barbara Hafner, and Richard Peto, "Mortality in Relation to Smoking: 22 Years' Observations on Female British Doctors," *British Medical Journal* 280, no. 6219 (1980): 967–971.
26. Royal College of Physicians. *Smoking and Health Now: A new report and summary on smoking and its effects on health*. London: Royal College of Physicians of London, Pitman Medical and Scientific Publishing Company, 1971.
27. Abraham M. Lilienfeld, "Epidemiology of Infectious and Non-Infectious Disease: Some Comparisons," *American Journal of Epidemiology* 97, no. 3 (1973): 135–147.
28. Hal Morgenstern, "Uses of Ecologic Analysis in Epidemiologic Research," *American Journal of Public Health* 72, no. 12 (1982): 1336–1344.
29. Leo A. Goodman, "Ecological Regressions and Behavior of Individuals," *American Sociological Review* 18 (1953): 663–664.
30. Steven Piantadosi, David P. Byar, and Sylvan B. Green, "The Ecological Fallacy," *American Journal of Epidemiology* 127, no. 5 (1988): 893–904.
31. William S. Robinson, "Ecological Correlations and the Behavior of Individuals," *International Journal of Epidemiology* 38, no. 2 (2009): 337–341.
32. H. C. Selvin, "Durkheim's Suicide and Problems of Empirical Research," *American Journal of Sociology* 63, no. 6 (1958): 607–619.

33. Centers for Disease Control and Prevention Epidemiology Program Office. "An Outbreak of Gastrointestinal Illness Following a Church Supper, Lycoming, Oswego County, New York, June 19, 1940." Atlanta, GA: US Department of Health and Human Services, Public Health Service, 1981.
34. Centers for Disease Control and Prevention. "Foodborne Diseases Active Surveillance Network (FoodNet)." Available at http://www.cdc.gov/foodnet/.
35. Centers for Disease Control and Prevention, "National Outbreak Reporting System (NORS, Food)." Available at http://www.cdc.gov/nors/.
36. Michael Gross, "Oswego County Revisited," *Public Health Reports* 91 (1976): 168–170.
37. Alfredo Morabia and Anne Hardy, "The Pioneering Use of a Questionnaire to Investigate a Food Borne Disease Outbreak in Early 20th Century Britain," *Journal of Epidemiology and Community Health* 59 (2005): 94–99.

Problem Set: Chapter 1

1. A large Coast Guard training center served a special breakfast for 535 recruits who were landlocked over a major holiday weekend. The clinic physician at the training center, Dr. Treadwater, began seeing recruits with the same symptoms (nausea, vomiting, and diarrhea) throughout the morning and into the afternoon. In total, 58 recruits were being treated for what appeared to be the same illness before the end of the day. Dr. Treadwater felt strongly that the clinic was dealing with a foodborne outbreak. Help Dr. Treadwater by calculating the crude attack rate for the suspected foodborne outbreak following the special breakfast.
2. Dr. Treadwater was concerned that additional recruits may have been ill but not ill enough to come to the clinic for treatment. The physician assistant at the clinic, PA Ondeck, was charged with surveying all of the recruits about which food items they ate at the special breakfast and whether or not they experienced similar symptoms. PA Ondeck found a few additional cases, and he created a line listing of the recruits' responses to his questions. The line listing appears below.

FOOD	CONSUMED FOOD		DID NOT CONSUME FOOD	
	NUMBER WELL	NUMBER ILL	NUMBER WELL	NUMBER ILL
Tomato juice	204	47	263	21
Cantaloupe	290	53	177	15
Creamed chipped beef	147	60	320	8
Potatoes	161	44	306	24
Eggs	169	39	298	29
Pastry	204	34	263	34
Toast	238	46	229	22
Milk	301	50	166	18

Help PA Ondeck by creating a table that includes the food-specific attack rates following the special breakfast that morning.

3. Dr. Treadwater asks PA Ondeck for the rate ratios to narrow down which food item might have been contaminated with the causative agent of the outbreak. Help PA Ondeck by calculating the rate ratios for each item to one decimal point and enter them into your table.
4. Dr. Treadwater asks you which food item you believe is the likely cause of this “common source” outbreak. Explain your choice.
5. Why might so many of the recruits who ate the creamed chipped beef not have become ill? Why might some who did not eat it still become ill?
6. What additional investigations could be done to determine the source of the likely causative agent?
7. How can such foodborne outbreaks be prevented at the Coast Guard training center in the future?

2 | Threads of Epidemiologic History

Don't you (forget about me)[1] . . .

Keith Forsey and Steve Schiff (1984)

Epidemiology is an eclectic discipline, so its history is generally interwoven with that of other academic disciplines. In the nineteenth century, however, epidemiology began developing its own philosophy, concepts, and methods. This chapter focuses on two major components that helped the discipline develop its unique framework: theories of disease etiology and the development of epidemiologic methods.

Theories of Disease Etiology

The idea that the environment can influence the occurrence of disease had its origins in antiquity. For instance, circa 400 BCE, Hippocrates' *On Airs, Waters, and Places* stressed:

> Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year . . . Then the winds . . . We must also consider the qualities of the waters . . . And the mode in which the inhabitants live, and what are their pursuits, whether they are fond of drinking and eating to excess, and given to indolence, or are fond of exercise and labor, not given to excess in eating and drinking.[2]

The Greek physician was not alone in his concerns. In the first century BCE, the Roman architect for Caesar Augustus, Vitruvius, wrote about the dangers of locating a city near fetid swamplands:

> For when the morning breezes blow toward the town at sunrise, if they bring with them mist from marshes and, mingled with the mist, the poisonous