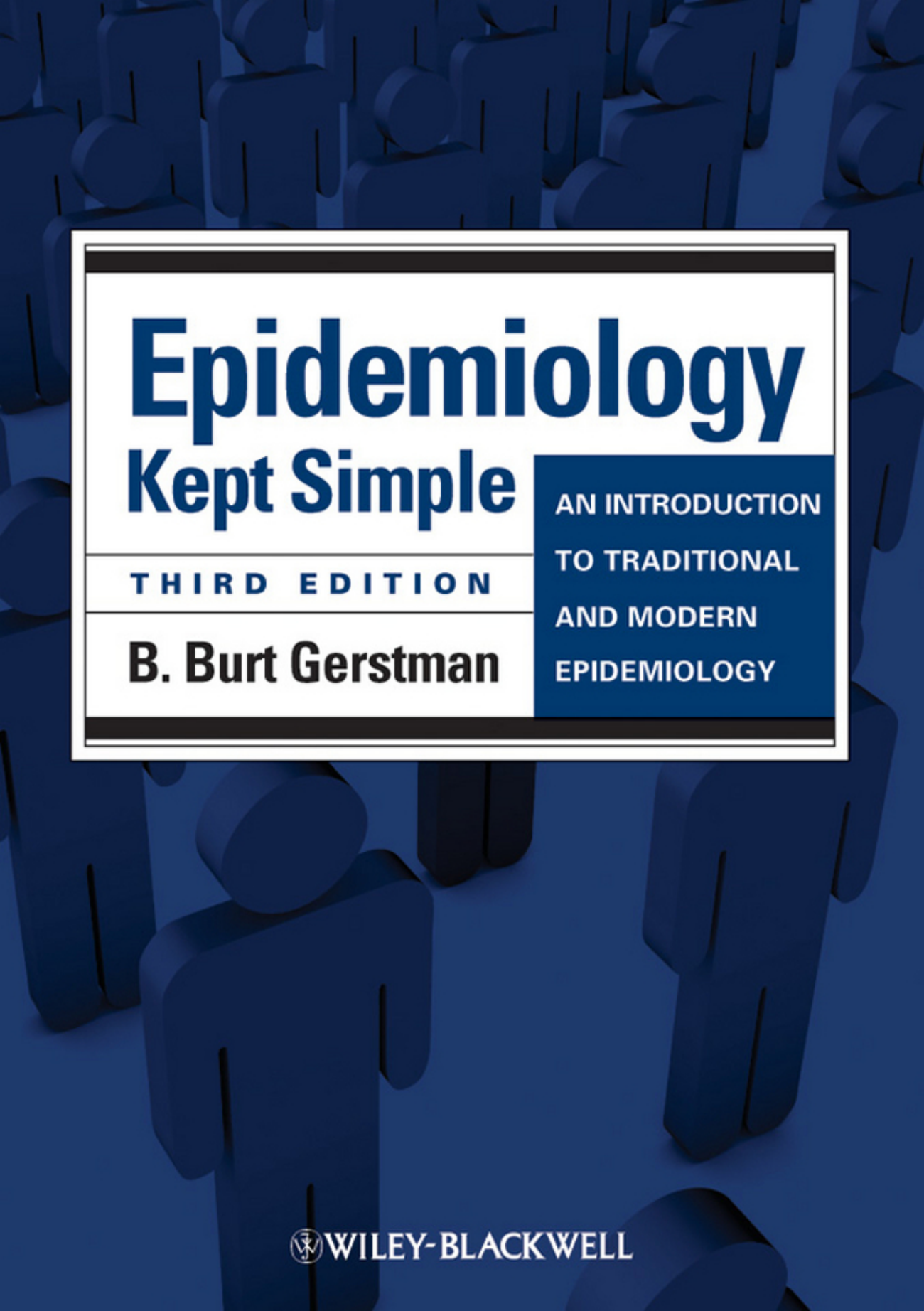

Epidemiology
Kept Simple
AN INTRODUCTION TO TRADITIONAL AND MODERN EPIDEMIOLOGY
THIRD EDITION
B. Burt Gerstman
WILEY-BLACKWELL

Epidemiology Kept Simple

An introduction to traditional and modern epidemiology

Epidemiology Kept Simple

An introduction to traditional and modern epidemiology

B. Burt Gerstman

Department of Health Science,
San José State University,
San José, CA,
USA,
www.wiley.com/go/gerstman
and
www.sjsu.edu/faculty/gerstman/eks

THIRD EDITION

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2013, © 2013 by John Wiley & Sons, Ltd.

Previous editions: 2003 by Wiley-Liss, Inc.

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

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 as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Gerstman, B. Burt.
Epidemiology kept simple : an introduction to traditional and modern epidemiology/
B. Burt Gerstman. –3rd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-4443-3608-5 (pbk.)
I. Title.
[DNLM: 1. Epidemiology. 2. Epidemiologic Methods. WA 105]
614.4–dc23
2012037293

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover image: iStock File #14279098.
Cover design by Modern Alchemy LLC.

Typeset in 9.5/12pt Meridien by Laserwords Private Limited, Chennai, India.

1 2013

To Linda

Contents

Preface to the Third Edition, xi

Preface to the First Edition, xiii

Acknowledgments, xv

1 Epidemiology Past and Present, 1
- 1.1 Epidemiology and its uses, 2
- 1.2 Evolving patterns of morbidity and mortality, 5
- 1.3 Selected historical figures and events, 8
- 1.4 Chapter summary, 30

Review questions, 31
References, 32

2 Causal Concepts, 36
- 2.1 Natural history of disease, 36
- 2.2 Variability in the expression of disease, 40
- 2.3 Causal models, 41
- 2.4 Causal inference, 48

Exercises, 58
Review questions, 61
References, 63

3 Epidemiologic Measures, 66
- 3.1 Measures of disease frequency, 67
- 3.2 Measures of association, 74
- 3.3 Measures of potential impact, 79
- 3.4 Rate adjustment, 82

Exercises, 90
Review questions, 98
References, 99
Addendum: additional mathematical details, 101

4 Descriptive Epidemiology, 104
- 4.1 Introduction, 104
- 4.2 Epidemiologic variables, 108
- 4.3 Ecological correlations, 116

Exercises, 121
Review questions, 123
References, 124

5 Introduction to Epidemiologic Study Design, 126
5.1 Etiologic research, 126
5.2 Ethical conduct of studies involving human subjects, 129
5.3 Selected study design elements, 130
5.4 Common types of epidemiologic studies, 137
Exercises, 138
Review questions, 140
References, 141

6 Experimental Studies, 142
6.1 Introduction, 142
6.2 Historical perspective, 144
6.3 General concepts, 146
6.4 Data analysis, 152
Exercises, 156
Review questions, 157
References, 157

7 Observational Cohort Studies, 159
7.1 Introduction, 159
7.2 Historical perspective, 161
7.3 Assembling and following a cohort, 163
7.4 Prospective, retrospective, and ambidirectional cohorts, 164
7.5 Addressing the potential for confounding, 165
7.6 Data analysis, 166
7.7 Historically important study: Wade Hampton Frost's birth cohorts, 170
Exercises, 174
Review questions, 177
References, 177

8 Case–Control Studies, 180
8.1 Introduction, 180
8.2 Identifying cases and controls, 182
8.3 Obtaining information on exposure, 185
8.4 Data analysis, 186
8.5 Statistical justifications of case–control odds ratio as relative risks, 193
Exercises, 194
Review questions, 198
References, 199

9 Error in Epidemiologic Research, 201
9.1 Introduction, 201
9.2 Random error (imprecision), 203
9.3 Systematic error (bias), 209
Exercises, 217
Review questions, 219
References, 220

10 Screening for Disease, 222
10.1 Introduction, 223
10.2 Reliability (agreement), 224
10.3 Validity, 228
Summary, 238
Exercises, 239
Review questions, 243
References, 243
10.4 Chapter addendum (case study), 244
Further reading—screening for HIV, 248
Further reading—general concepts of screening, 248
Answers to case study: screening for antibodies to the human immunodeficiency virus, 249

11 The Infectious Disease Process, 255
11.1 The infectious disease process, 255
11.2 Herd immunity, 265
Exercises, 267
Review questions, 268
References, 270

12 Outbreak Investigation, 271
12.1 Background, 272
12.2 CDC prescribed investigatory steps, 273
Review questions, 282
References, 283
References—a drug–disease outbreak, 286

13 Confidence Intervals and *p*-Values, 302
13.1 Introduction, 303
13.2 Confidence intervals, 304
13.3 *p*-Values, 312
13.4 Minimum Bayes factors, 319
References, 322

14 Mantel–Haenszel Methods, 323
14.1 Ways to prevent confounding, 323
14.2 Simpson's paradox, 325
14.3 Mantel–Haenszel methods for risk ratios, 325
14.4 Mantel–Haenszel methods for other measures of association, 329
Exercise, 335
References, 335

15 Statistical Interaction: Effect Measure Modification, 337
15.1 Two types of interaction, 337
15.2 Chi-square test for statistical, 340
15.3 Strategy for stratified analysis, 342

Exercises, 344
References, 345

16 Case Definitions and Disease Classification, 347
16.1 Case definitions, 347
16.2 International classification of disease, 351
16.3 Artifactual fluctuations in reported rates, 353
16.4 Summary, 354
References, 355

17 Survival Analysis, 356
17.1 Introduction, 356
17.2 Stratifying rates by follow-up time, 359
17.3 Actuarial method of survival analysis, 360
17.4 Kaplan–Meier method of survival analysis, 362
17.5 Comparing the survival experience of two groups, 364
Exercises, 369
References, 371

18 Current Life Tables, 373
18.1 Introduction, 373
18.2 Complete life table, 374
18.3 Abridged life table, 380
Exercises, 383
References, 384

19 Random Distribution of Cases in Time and Space, 385
19.1 Introduction, 385
19.2 The Poisson distribution, 386
19.3 Goodness of fit of the Poisson distribution, 390
19.4 Summary, 394
Exercises, 395
References, 396

Answers to Exercises and Review Questions, 398
Appendix 1: 95% Confidence Limits for Poisson Counts, 434
Appendix 2: Tail Areas in the Standard Normal (Z) Distribution: Double These Areas for Two-Sided p-Values, 436
Appendix 3: Right-Tail Areas in Chi-Square Distributions, 439
Appendix 4: Case Study—Cigarette Smoking and Lung Cancer, 441
Appendix 5: Case Study—Tampons and Toxic Shock Syndrome, 448

Index, 455

Preface to the Third Edition

This major re-write of *Epidemiology Kept Simple* was pursued with the following objectives in mind:

- To address the American Schools of Public Health (ASPH) Epidemiology Competencies[a] in the first dozen chapters of the book.
- To introduce epidemiologic measures early in the book's progression so that they can be used throughout.
- To devote full chapters to the following topics: Descriptive Epidemiology (Chapter 4), Epidemiologic Study Design (Chapter 5), Experimental Studies (Chapter 6), Observational Cohort Studies (Chapter 7), and Case–control Studies (Chapter 8).
- To provide more frequent Illustrative Examples.
- To provide additional exercises and review questions to help students learn the material.
- To extend the section on epidemiologic history (Section 1.3) to address developments in the first half of the 20th century.

The ASPH Epidemiology Competencies alluded to in the first bullet are: ''Upon graduation a student with an MPH (Master of Public Health) should be able to:

1 Explain the importance of epidemiology for informing scientific, ethical, economic, and political discussion of health issues.
2 Describe a public health problem in terms of magnitude, person, time, and place.
3 Apply the basic terminology and definitions of epidemiology.
4 Identify key sources of data for epidemiologic purposes.
5 Calculate basic epidemiology measures.
6 Evaluate the strengths and limitations of epidemiologic reports.
7 Draw appropriate inferences from epidemiologic data.
8 Communicate epidemiologic information to lay and professional audiences.
9 Comprehend basic ethical and legal principles pertaining to the collection, maintenance, use, and dissemination of epidemiologic data.
10 Identify the principles and limitations of public health screening programs.

This list provides a framework for approaching the instruction of introductory epidemiology to MPH students and, in my opinion, to undergraduates as well.[b] Most of these Competencies draw from several areas of epidemiology. For example, Competency 1 (Explain the importance of epidemiology for ...) requires an understanding

[a] Calhoun, J.G., Ramiah, K., Weist, E.M., and Shortell, S.M. (2008). Development of a core competency model for the master of public health degree. *American Journal of Public Health*, **98** (9), 1598–1607.
[b] The ASPH maintains separate learning outcomes for undergraduates—see *http://www.aacu.org/public_health/documents/Recommendations_for_Undergraduate_Public_Health_Education.pdf* for details.

of epidemiologic history (Chapter 1), the causes and prevention of disease (Chapter 2), descriptive epidemiology (Chapter 4), analytic epidemiology (Chapters 5–8), and so on. Therefore, I chose not to cover competencies on a one competency per chapter basis.[c] However, by covering the first dozen chapters in this book the student will be well on the road to achieving all ten competencies.

The ASPH competency list is intended to improve individual performance and enhance communication and coordination across courses and programs. In that sense, it serves a useful purpose. However, I believe it is important to view this list as a starting point and not as an ultimate destination. To view the discipline of epidemiology (taught at any level) as a list of competencies does not adequately acknowledge the discipline's depth, breadth, and complexity. To achieve this deeper understanding and appreciation of epidemiologic research and practice requires diligence, discipline, constant questioning, experience, a drive from *within*, and a healthy dose of epistemological modesty. After three decades in the profession, I still cannot say that I've fully mastered every epidemiologic competency. However, one must continue to push the boulder up hill. As Camus has said, ''One must imagine Sisyphus happy.'' So let the struggle begin.

May all your rates be adjusted, your estimates precise, and your inferences unbiased,

B. Burt Gerstman
Aptos, California

[c] Three of the Competencies are specific and are therefore covered in single chapters. Chapters 5 (Descriptive Epidemiology) corresponds well to Competency 2, Chapter 3 (Epidemiologic Measures) corresponds to Competency 5, and Chapter 10 (Screening for Disease) corresponds with Competency 10.

Preface to the First Edition

> Things should be made as simple as possible, but not any simpler.
>
> *Albert Einstein*

Who studies epidemiology and why they bother

What is epidemiology?

Epidemiology studies the causes, transmission, incidence, and prevalence of health and disease in human populations. Medical and public health disciplines use epidemiologic study results to solve and control human health problems.

Who studies epidemiology?

Traditionally, epidemiology has been studied as a core science of public health. As such, it provided the objective basis for disease prevention and health promotion. Public health professionals of all types must communicate risk and read epidemiologic information. Epidemiology provides the tools to evaluate health problems and policies on a population basis. Epidemiology is also included in many undergraduate and graduate programs in medicine, the allied health professions, community health, environmental health, occupational health and industrial hygiene, health education, and health services administration. Because of its applicability and utility, epidemiology continues to gain a still wider audience.

Epidemiology as a liberal art

The study of epidemiology also belongs in the liberal arts. A liberal arts education provides general knowledge and develops overall intellectual capacities. Epidemiology fits nicely into an undergraduate liberal arts course of study because (Fraser, 1987):

- It uses the scientific method.
- It develops and improves one's ability to reason inductively (reasoning from the specific to the general).
- It develops and improves one's ability to reason deductively (logical conclusion that follows from a premise).
- It develops and improves one's ability to reason by analogy.
- It develops one's concern for aesthetic values (appreciation of elegance, beauty, simplicity, grace).

- It emphasizes investigative method rather than arcane knowledge and specialized investigative tools.

Moreover, epidemiologists benefit from studying the humanities. By studying the humanities, epidemiologists learn who they are, what is right, and how to think and act. Studying the humanities encourages epidemiologists to focus their skills on the people they serve while increasing flexibility of perspective, encouraging nondogmatisms, improving critical thinking skills, and promoting a better balance of values and ethics (Weed, 1995).

Other reasons to study epidemiology

There are still other reasons to study epidemiology. One such reason is to better understand the mounting epidemiologic information we receive on a regular basis. Much of this information is confusing and some of it is apparently contradictory. To effectively use epidemiologic information, we must understand its basis, its strengths, and its limitations. Without understanding the basis of epidemiologic research, we cannot make informed health decisions for ourselves and others.

Moreover, as involved citizens and voters, we often need to evaluate potential risks and benefits of public and private interventions and policies. For example, we may be called upon to vote on regulations to allow the construction of an industrial facility in our community. To make an informed decision, we must compare the potential economic benefits of the development to the potential environmental hazards it might present. Issues like this respond to epidemiologic analysis by preparing us to weigh the risks and benefits of an intervention on a population basis.

Finally, today's job market seeks people with epidemiologic competencies, such as those associated with data collection, risk/benefit analysis, survey methodology, and outcomes evaluation. These epidemiologic job skills might be useful in your current job and are transferable to other jobs as well.

And, yes, there is another reason to study epidemiology: it is inherently interesting. The challenges of disease detectives have captured the public's interest, as I hope this book will capture yours.

B. Burt Gerstman
San José, California

References

Fraser, D.W. (1987) Epidemiology as a liberal art. *New England Journal of Medicine*, **316**, 309–314.
Weed, D.L. (1995). Epidemiology, the humanities, and Public Health. *American Journal of Public Health*, **85**, 914–918.

Acknowledgments

The classical Hippocratic oath pledges first to honor one's teachers. This, too, I shall do. I am indebted to my teachers of epidemiology, for without whom this book would not have been possible. These include Lawrence Glickman, who first introduced me to epidemiology when I was a student at Cornell, and Maurice ''Pete'' White, ''who implanted in me the idea that clinical medicine should be by hard facts and not supposition. I am also indebted to the teachers I met at UC Berkeley and UC Davis, including Bill Reeves, Nick Jewell, Warren Winkelstein, Mary Claire-King, Allan Smith, Len Syme, Aron Antonovsky, James Beaumont, Steve Samuels, Jessica Utts, and Hari Mukerjee.

I owe a great deal to the mentors that I met as part of the Public Health Service Training Program in Epidemiology, notably Joyce Piper and Frank Lundin. Although Joyce passed away too young, her intelligence and spirit remain inspirational. I am grateful to Frank for teaching me about the essential nature of longitudinal analysis as risks vary over time.

I would also like to honor selected published works that were inspirational and helpful to me when I was learning epidemiology. Notable in this regard are *Epidemiologic Research* by Kleinbaum, Kupper, and Morgenstern (1982, Van Nostrand Reinhold), the published articles by Olli Miettinen (especially those from the 1970s and 1980s), *Foundations of Epidemiology* by Lilienfeld and Lilienfeld (1980, Oxford University Press), Alvan Feinstein's instructional articles on clinical epidemiology published in the 1960s and 1970s, and Ken Rothman's *Modern Epidemiology* (1986, Little, Brown and Company). More advanced studies of epidemiology are available through these sources.

I would also like to acknowledge Joe Abramson for creating and supporting the incredibly complete and well-documented epidemiologic computer application *WinPEPI*.[a] In a similar vein, I am grateful to the team of Dean, Sullivan, and Soe for their website of *www.OpenEpi.com*. I would also like to thank Andy Dean, Kevin Sullivan, and Mark Myatt for their helpful suggestions in preparation of earlier editions of this book, and Ken Rothman for his comments in the preparation of the Second Edition.

I have had a number of different assistants on this project, all of whom deserve praise. Jean Shiota of San Jose State University's Instructional Resource Center has prepared or assisted in the preparation of the illustrations for all three editions of the book. My student research assistants have included Karen Garcida-Ankele (First Edition), Dana Zive-Ducker and Elizabeth Edwards (Second Edition), and Deborah Danielewicz (current Edition).

[a] Download from *http: www.brixtonhealth.com*; see *www.epi-perspectives.com/content/8/1/1* for details.

Provision of general career and administrative support is often overlooked on projects like this. Let me not be negligent in this regard by recognizing the early career support of Lynn Bosco, Chuck Anello, Joel Kuritsky, and Jerry Faich (U.S. Food and Drug Administration) and academic support of Dean Rose Tseng, Dean Michael Ego, and Dean Charles Bullock, and Chairs Bill Washington and Kathleen Roe (San Jose State University).

Finally, I express my deepest appreciation to my family, and especially to my wife M. Linda Gerstman, who has always given me the time, support, and encouragement to take on large challenges such as the writing of this book. Thanks Linda. You're the greatest.

CHAPTER 1
Epidemiology Past and Present

1.1 Epidemiology and its uses
- What is epidemiology?
- What is public health?
- What is health?
- Additional useful terms
- Uses of epidemiology

1.2 Evolving patterns of morbidity and mortality
- Twentieth century changes in demographics and disease patterns
- Mortality trends since 1950
- Trends in life expectancy

1.3 Selected historical figures and events
- Roots of epidemiology
- John Graunt
- Germ theory
- Médecine d'observation and La Méthode Numerique (Pinel and Louis)
- The London Epidemiological Society
- William Farr
- John Snow
  - Cholera in Victorian England
  - Miasma theory of transmission
  - Snow's theory
  - Snow's ecological analysis
  - Snow's retrospective cohort analysis
  - Snow's case series
  - Publication
- Twentieth-century epidemiology
- Emile Durkheim
- Joseph Goldberger
- The British Doctors Study

1.4 Chapter summary
- Epidemiology and its uses
- Evolving patterns of morbidity and mortality
- Selected historical figures and events

Review questions

References

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

1.1 Epidemiology and its uses

What is epidemiology?

The word **epidemiology** is based on the Greek roots *epi* (upon), *demos* (the people, as in ''democracy'' and ''demography''), and *logia* (''speaking of,'' ''the study of''). Specific use of the term in the English language dates to the mid-19th century (*Oxford English Dictionary*), around the time the London Epidemiological Society was founded in 1850. Since then, *epidemiology* has defined itself in many ways, including:

- the study of the distribution and determinants of diseases and injuries in populations (Mausner and Baum, 1974);
- the study of the occurrence of illness (Gaylord Anderson cited in Cole, 1979, p. 15);
- a method of reasoning about disease that deals with biological inferences derived from observations of disease phenomena in population groups (Lilienfeld, 1978b, p. 89);
- the quantitative analysis of the circumstances under which disease processes, including trauma, occur in population groups, and factors affecting their incidence, distribution, and host responses, and the use of this knowledge in prevention and control (Evans, 1979, p. 381).

A widely accepted contemporary definition of epidemiology identifies the discipline as ''the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems'' (Last, 2001).

The word *epidemiology* is, of course, based on the word **epidemic**. This term dates back to the time of Hippocrates, circa 400 BCE. Until not too long ago, *epidemic* referred only to the rapid and extensive spread of an infectious disease within a population. Now, however, the term applies to any health-related condition that occurs in clear excess of normal expectancy. For example, one may hear mention of an ''epidemic of teen pregnancy'' or an ''epidemic of violence.'' This broader use of the term reflects epidemiology's expansion into areas beyond infectious disease control to include the study of health and health-related determinants in general. In this non-limiting sense, epidemiology is still the study of epidemics and their prevention (Kuller, 1991).

In addition, epidemiology is becoming increasingly integrated in biomedical research and health care. Note, however, that the main distinction between *epidemiology* and *clinical medicine* is their primary **unit of concern**. The primary unit of concern for the epidemiologist is ''an aggregate of human beings'' (Greenwood, 1935). Compare this with clinical medicine, whose main unit of concern is the individual. A metaphor that compares epidemiology with clinical medicine discusses a torrential storm that causes a break in the levees. People are being washed away in record numbers. Under such circumstances, the physician's task is to offer lifejackets to people one at a time. In contrast, the epidemiologist's task is to stem the tide of the flood to mitigate the problem and prevent future occurrences.

What is public health?

Like epidemiology, public health has been defined in many different ways including ''organized community effort to prevent disease and promote health (Institute of

Medicine, 1988) and ''one of the efforts organized by society to protect, promote, and restore the people's health (Last 2001). By any definition, the aim of public health is to reduce injury, disability, disease, and premature death in the population. Public health is thus a mission comprising many activities, including but not limited to epidemiology. Epidemiology is a ''study of'' with many applications, while public health is an undertaking.

Note that epidemiology is one of the core disciplines of public health. Other core disciplines in public health include biostatistics, environmental health sciences, health policy and management, and social and behavioral sciences (Calhoun *et al*., 2008). The practice of public health also requires cross-cutting interdisciplinary competencies in areas such as communication, informatics, culture and diversity, and public health biology.

What is health?

Health itself is not easily defined. The standard medical definition of health is ''the absence of disease.'' Dis-ease, literally the absence of ''ease,'' is when something is wrong with a bodily or mental function. The World Health Organization in the preamble to its 1948 constitution defined health as ''a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.''

Walt Whitman (1954, p. 513), in his poetic way, defined health as:

> the condition [in which] the whole body is elevated to a state by other unknown—inwardly and outwardly illuminated, purified, made solid, strong, yet buoyant. A singular charm, more than beauty, flickers out of, and over, the face—a curious transparency beams in the eyes, both in the iris and the white—temper partakes also. The play of the body in motion takes a previously unknown grace. Merely to move is then a happiness, a pleasure—to breathe, to see, is also. All the before hand gratifications, drink, spirits, coffee, grease, stimulants, mixtures, late hours, luxuries, deeds of the night seem as vexatious dreams, and now the awakening; many fall into their natural places, wholesome, conveying diviner joys.

This passage from Whitman address *quality of life*, an area of increasing interest to epidemiologists.

Additional useful terms

One of the ten American Schools of Public Health MPH Epidemiology competencies is to ''apply the basic terminology and definitions of epidemiology'' (Calhoun *et al*., 2008). Therefore, terminology will be introduced throughout this book. Table 1.1 lists definitions for several standard terms. For example, an **epidemic** is the occurrences of disease in clear excess of normalcy, while a **pandemic** is an epidemic that affects several countries or continents. An **endemic** disease is one that is consistently present in the environment. The term endemic is also used to refer to a normal or usual rate of disease. An excellent source for epidemiologic definitions is *The Dictionary of Epidemiology* (Porta, 2008), which is updated periodically.

Some terms used in the field are not readily defined in a singular way. For example, some sources differentiate between disease, illness, and sickness. Susser (1973) defines

Table 1.1 Selected terms briefly defined.

Epidemiology: the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems
Public health: organized effort to prevent disease and promote health
Endemic: occurring at a consistent or regular rate
Epidemic: occurring in clear excess of normalcy
Pandemic: an epidemic that affects several countries or continents
Morbidity: related to or caused by disease or disability
Mortality: related to death

disease as the medically applied term for a physiological or psychological dysfunction; **illness** is what the patient experiences; and **sickness** is the state of dysfunction of the social role of an ill person. In contrast, one source considers ''disease'' a subtype of ''illness'' (Miettinen and Flegel, 2003). While yet in other contexts, ''disease'' is merely a general term used to refer to any health-related outcome or condition. Thus, the use of epidemiologic terminology is context specific and is, at times, controversial.

Uses of epidemiology

Epidemiologic practice is characterized by a close connection between the scientific study of the causes of disease, and the application of this knowledge to treatment and prevention (especially the later). The discipline covers a broad range of activities, including conducting biomedical research, communicating research findings, and participating with other disciplines and sectors in deciding on public health practices and interventions.

A sample of epidemiology's varied concerns include studies of the effects of environmental and industrial hazards, studies of the safety and efficacy of medicines and medical procedures, studies of maternal and child health, studies of food safety and nutrition, studies of the long-term effects of diet and lifestyle, surveillance and

Table 1.2 General uses of epidemiology (Morris, 1957).

1	In **historical study** of the health of the community and of the rise and fall of diseases in the population; useful ''projections'' into the future may also be possible.
2	For **community diagnosis** of the presence, nature, and distribution of health and disease among the population, and the dimensions of these in incidence, prevalence, and mortality; taking into account that society and health problems are changing.
3	To study the **workings of health services**. This begins with the determination of needs and resources, proceeds to analysis of services in action and, finally, attempts to appraise. Such studies can be comparative between various populations.
4	To estimate, from the common experience, the **individual's chances and risks of disease**.
5	To help **complete the clinical picture**: by including all types of cases in proportion; by relating clinical disease to subclinical; by observing secular changes in the character of disease, and its picture in other countries.
6	In **identifying syndromes** from the distribution of clinical phenomena among sections of the population.
7	In the **search for causes of health and disease**, starting with the discovery of groups with high and low rates, studying these differences in relation to differences in ways of living; and, where possible, testing these notions in actual practice among populations.

control of communicable and noncommunicable diseases, ascertainment of personal and social determinants of health and ill-health, medico-legal attribution of risk and responsibility, screening and early detection of the population for disease, and the study of health-care services. Because findings from epidemiologic investigations are linked to health policy, epidemiologic studies often have important legal, financial, and political consequences.

More than half a century ago, Morris (1957) described seven uses of epidemiology. These seven uses, listed in Table 1.2, have stood the test of time. The seventh use, search for causes, is perhaps the most important current application because of its essential role in effective disease prevention.

1.2 Evolving patterns of morbidity and mortality

Twentieth century changes in demographics and disease patterns

The theory of **epidemiologic transition** focuses on the dramatic changes in morbidity and mortality that have occurred in relation to demographic, biologic, and socioeconomic factors during the 20th century (Omran, 1971). Ample evidence exists to document a transition from infectious diseases as the predominant causes of morbidity and mortality to a predominance of noninfectious diseases (Table 1.3). The transition from predominantly infectious to noninfectious causes resulted from changes in society at large and improvements in medical technology. Steady economic development led to better living conditions, improved nutrition, decreases in childhood mortality, diminished fertility rates, and technological advances in medicine.

Decreases in mortality and fertility led to a substantial shift in the age distribution of populations, especially in industrialized societies, a phenomenon known as the **demographic transition** (Figure 1.1). With this now familiar demographic shift

Table 1.3 Leading causes of death in the United States, 1900 and 2007.[a]

Rank	1900[b]	2007[c]
1.	Pneumonia (all forms) and influenza [202.2]	Diseases of the heart [204.3]
2.	Tuberculosis (all forms) [194.4]	Malignant neoplasms (cancers) [186.6]
3.	Diarrhea, enteritis, and ulceration of the intestines [142.7]	Cerebrovascular diseases (stroke) [45.1]
4.	Diseases of the heart [137.4]	Chronic lower respiratory diseases [42.4]
5.	Intracranial lesions of vascular origin [106.9]	Accidents (unintentional injuries) [41.0]
6.	Nephritis (all forms) [88.6]	Alzheimer's disease [24.7]
7.	All accidents [72.3]	Diabetes mellitus (diabetes) [23.7]
8.	Cancer and other malignant tumors [64.0]	Influenza and pneumonia [17.5]
9.	Senility [50.2]	Nephritis, nephrotic syndrome and nephrosis (kidney diseases) [15.4]
10.	Diphtheria [40.3]	Septicemia [11.5]

[a] Crude death rates per 100 000 are listed in square brackets. Rates have not been adjusted for age differences in the population and, therefore, should not be compared between time periods.
[b] Source: National Office of Vital Statistics, 1947.
[c] Source: Xu *et al*., 2010.

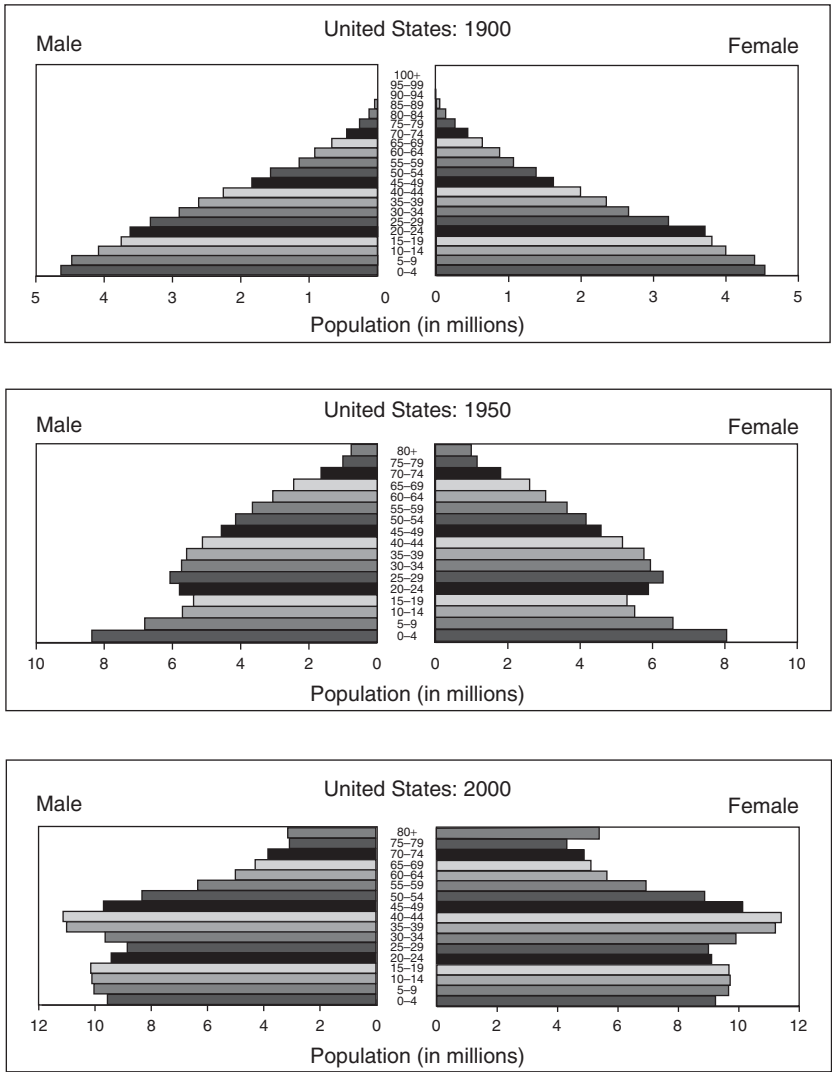


Figure 1.1 Population pyramids for the United States, 1900, 1950, and 2000 (Sources: Bureau of the Census, 1904; U.S. Census Bureau International Data Base, 2002).

came a concomitant rise in age-related diseases such as atherosclerotic cardiovascular and cerebrovascular disease, cancer, chronic lung disease, diabetes and other metabolic diseases, liver disease, musculoskeletal disorders, and neurological disorders. Many of these noncontagious diseases are thought to have important lifestyle components rooted in behaviors such as smoking, dietary excesses, and physical inactivity (''diseases of civilization''). As of the mid-20th century, these prevalent chronic diseases were viewed primarily as an intrinsic property of aging (so-called degenerative diseases). Now, however, these diseases are regarded as a diverse group of pathologies with varied and complex etiologies. What brings them together as a

Table 1.4 Chronic diseases and their relation to selected, modifiable risk factors: + = established risk factor and ± = possible risk factor.

Cause	Cardiovascular disease	Cancer	Chronic lung disease	Diabetes	Cirrhosis	Musculoskeletal diseases	Neurologic disorders
Tobacco use	+	+	+			+	±
Alcohol use	±	+			+	+	+
High cholesterol	+						
High blood pressure	+						
Diet	+	+	±	±		+	±
Physical inactivity	+	+		+		+	
Obesity	+	+		+		+	+
Stress	±	±					
Environ. tobacco smoke	±	±	+				
Occupation		+	+		±	+	±
Pollution		+	+				+
Low socioeconomic status	+	+	+	+	+	+	

Based on Brownson *et al*. (1993, p. 4).

group is their insidious onset, long duration, and the fact that they seldom resolve spontaneously.

By the middle of the 20th century, epidemiologists came to realize that the limited tools they had developed to address acute infectious diseases were no longer sufficient in studying chronic ailments. Out of this awareness arose development of new investigatory tools—field surveys, cohort studies, case–control studies, and clinical trial—as will be addressed later in this book. Using these newly developed methods, epidemiologists identified risk factors that influence the incidence of many chronic conditions (Table 1.4).

Mortality trends since 1950

Figure 1.2 displays age-adjusted mortality rates for all causes combined and the six leading causes of death in the United States in 2006 for the years 1950 through 2006. Rates are plotted on a logarithmic scale, so even modest downward slopes represent large changes in the rates of occurrence. During this period, age-adjusted mortality for all causes combined decreased from 1446.0 per 100 000 in 1950 to 776.5 per 100 000, a 47% decline. An important component of this decline came from advances in preventing cardiovascular and cerebrovascular mortality. In 1950, mortality from heart disease occurred at the adjusted rate of 588.8 per 100 000. By 1992, this rate was cut by two-thirds, to 200.2 per 100 000.

Trends in life expectancy

Life expectancy is the average number of years of life a person is expected to live if current mortality rates in the population were to remain constant. In 1900, life

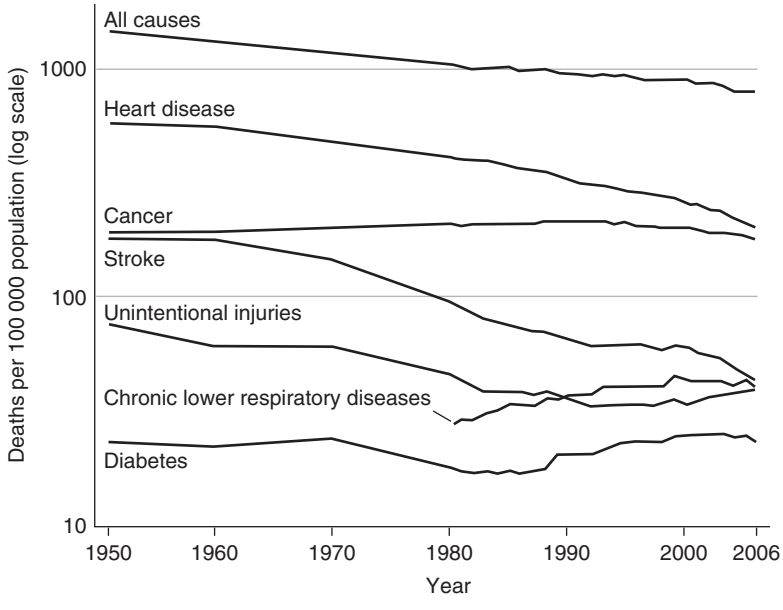


Figure 1.2 Age-adjusted death rates from 1950 to 2006, the United States, for the six leading causes of death in 2006 (Source: CDC/NCHS 2010).

expectancy at birth in the United States was 47.3 years. By 2006, life expectancy was 77.7 years (75.1 years for men and 80.2 years for women). Figure 1.3 charts this dramatic progress.

During the early part of the 20th century, increases in life expectancy can be traced to decreases in mortality at younger ages due primarily to improved sanitization and hygiene, improved nutrition, smaller family size, better provision of uncontaminated water, control of infectious disease vectors, pasteurization of milk, better infant and child care, and immunization (Doll, 1992). Since the middle of the century 20th century, life expectancy at older ages has shown significant increases. In 1950, a 65 year old man had a life expectancy of 12.8 remaining years; by 2000 this has increased to 16.0 years; by 2006 this had increased to 17.0 years (CDC/NCHS, 2010). For women, comparable increases have occurred. These increases can be traced to technological improvements in medical care (e.g., antibiotics, improvements in the safety of surgery, treatment of hypertension, etc.), dietary changes, avoidance of smoking, reductions in vascular diseases, and the pharmacologic control of high blood pressure and hyperlipidemia (Doll, 1992).

1.3 Selected historical figures and events

> A knowledge of epidemiological history, combined with a firm grasp of the statistical method were as essential parts of the outfit of the investigator in the field as was a grounding in bacteriology.
>
> *Major Greenwood*

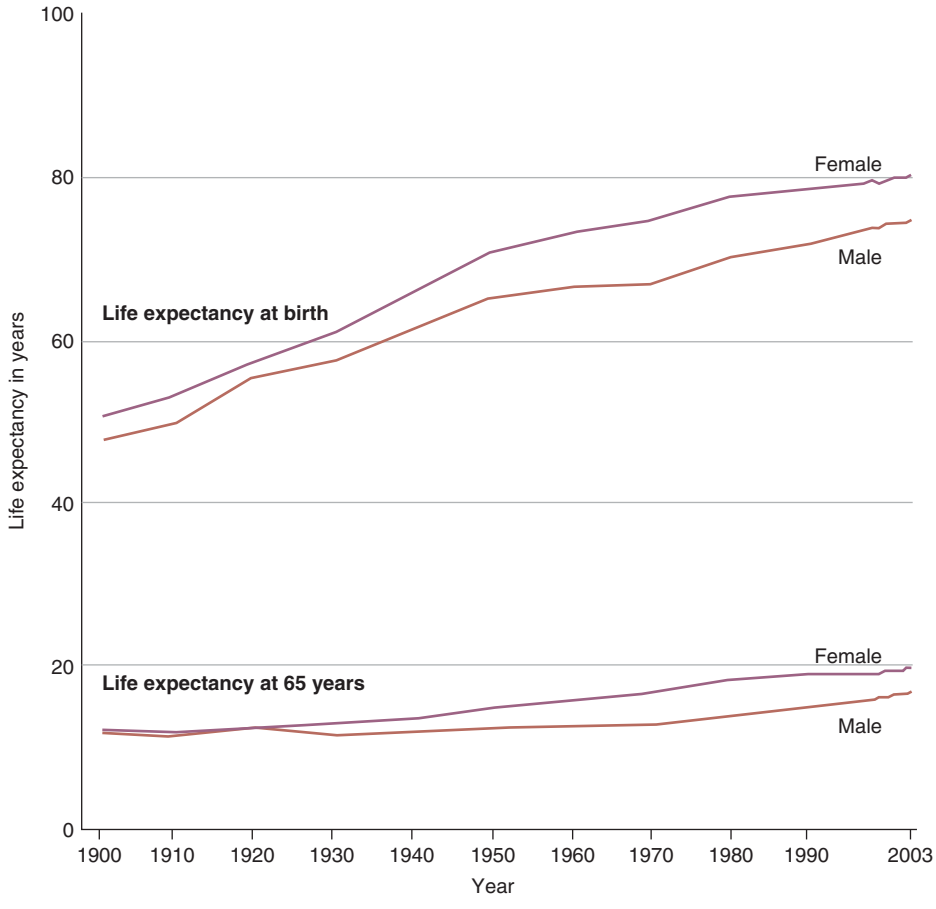


Figure 1.3 Life expectancy at birth and at age 65 by sex, United States, 1900–2003 (Source: CDC/NCHS, 2006).

Roots of epidemiology

Epidemiological insights into health and disease are probably as old as civilization itself. The Old Testament refers to the benefits of certain diets, the Greeks linked febrile illnesses to environmental conditions (''marsh fever''), and the Romans recognized the toxic effects of consuming wine from lead-glazed pottery.

Hippocrates (circa 460–388 BCE) is said to have prepared the groundwork for the scientific study of disease by freeing the practice of medicine from the constraints of philosophical speculation, superstition, and religion, while stressing the importance of careful observation in identifying natural factors that influenced health. In *Air, Waters, and Places* (Table 1.5), Hippocrates refers to environmental, dietary, behavioral, and constitutional determinants of disease. ''From these things, we must proceed to investigate everything else.'' Elsewhere, Hippocrates provides accurate descriptions of various clinical ailments, including tetanus, typhus, and tuberculosis.

Table 1.5 Part I of *On Air, Waters, and Places* (Hippocrates, 400 BCE).

Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces for they are not at all alike, but differ much from themselves in regard to their changes. Then the winds, the hot and the cold, especially such as are common to all countries, and then such as are peculiar to each locality. We must also consider the qualities of the waters, for as they differ from one another in taste and weight, so also do they differ much in their qualities. In the same manner, when one comes into a city to which he is a stranger, he ought to consider its situation, how it lies as to the winds and the rising of the sun; for its influence is not the same whether it lies to the north or the south, to the rising or to the setting sun. These things one ought to consider most attentively, and concerning the waters which the inhabitants use, whether they be marshy and soft, or hard, and running from elevated and rocky situations, and then if saltish and unfit for cooking; and the ground, whether it be naked and deficient in water, or wooded and well watered, and whether it lies in a hollow, confined situation, or is elevated and cold; 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, and not given to excess in eating and drinking.

A long period of relative quiescence in scientific medicine followed the Hippocratic era. In the 17th century scientific observation in medicine began to reawaken, dawning an upcoming Age of Enlightenment in the 18th century. This period is credited with the development of scientific methods based on systematized observation, experimentation, measurement, and a multistep process that advanced from theory to conclusion by testing and revising causal hypotheses. In summarizing the profound impact brought about by these changes, Ariel and Will Durant (1961, p. 601) wrote:

> Science now began to liberate itself from the placenta of its mother, philosophy. It developed its own distinctive methods, and looked to improve the life of man on the earth. This movement belonged to the heart of the Age of Reason, but it did not put its faith in ''pure reason''—reason independent of experience and experiment. Reason, as well as tradition and authority was now to be checked by the study and record of lowly facts; and whatever logic might say, science would aspire to accept only what could be quantitatively measured, mathematically expressed, and experimentally proved.

The features of scientific work—measuring, sequencing, classifying, grouping, confirming, observing, formulating, questioning, identifying, generalizing, experimenting, modeling, and testing—now took prominence.

A very early reawakening came with the work of the ''English Hippocrates'' **Thomas Sydenham** (1624–1689). Like Hippocrates, Sydenham stressed the need for careful observation for the advancement of health care. Using information combed from patients' records, Sydenham wrote about the prevalent diseases of his day. In a similar vein, Sydenham's contemporary **Bernardino Ramazzini** (1633–1714) published his comprehensive work *The Diseases of Workers* (*De Morbis Artificum Diatriba*). *The Diseases of Workers* discussed the hazards of various environmental irritants (chemicals, dust, metals, and abrasive agents) encountered in 52 different occupations. Renowned as an early expositor of specificity in linking environment cause to disease, Ramazzini set the stage for occupational medicine and environmental epidemiology. Not long after Ramazzini, the Englishman **Percival Pott** (1713–1788) identified chimney soot as the cause of enormously elevated rates of scrotal cancer in chimney sweeps (Pott,

1775/1790). This may have been the first link demonstrating a causal association between a malignancy and an environmental carcinogen.

John Graunt

The development of systems to collect the causes of death on a population basis was key to the development of epidemiology. The earliest tallying of deaths dates back to the reign of the Black Death (bubonic plague), when in the 14th and 15th centuries officials in Florence and Venice began keeping records of the number of persons dying, specifying cause of death in broad terms, such as plague/not plague (Saracci, 2001).

In England, the collection of death certificates began in selected parishes in 1592. However, it was not until the middle of the 17th century that this resource started to be used in an epidemiologic way by an intellectually curious London haberdasher by the name of **John Graunt** (1620–1674; Figure 1.4). Graunt tallied mortality

Figure 1.4 John Graunt (1620–1674).

statistics and made many forward-looking and insightful interpretations based these tallies in his publication *Natural And Political Observations Mentioned In A Following Index And Made Upon The Bills Of Mortality* (1662). Among his many observations, Graunt noted regional differences in mortality, high mortality in children (one-third of the population died before the age of 5), and greater mortality in men than women despite higher rates of physician visits in women (a phenomenon that still exists today). He noted that more boys than girls were born, debunked inflated estimates of London's population size, noted that population growth in London was due mostly to immigration, determined that plague claimed more deaths than originally thought, and documented an epidemic of rickets.

By starting with a hypothetical group of 100 people, Graunt constructed one of the first known life tables as follows. Out of 100 people born, Graunt projected the following expectations for survival (O'Donnell, 1936):

At the end of 6 years	64 of the initial 100 would be alive
At the end of 16 years	40 of the initial 100 would be alive
At the end of 26 years	25 of the initial 100 would be alive
At the end of 36 years	16 of the initial 100 would be alive
At the end of 46 years	10 of the initial 100 would be alive
At the end of 56 years	6 of the initial 100 would be alive
At the end of 60 years	3 of the initial 100 would be alive
At the end of 76 years	1 of the initial 100 would be alive
At the end of 80 years	0 of the initial 100 would be alive

Graunt recognized the importance of systematized record collection, was fastidious in his concern for accuracy, and took great care in scrutinizing the origins of data while being aware that certain forms of death tended to be misclassified. Given the period in which he lived and the limitations of its data, these are remarkable insights. It is therefore not surprising that many modern epidemiologists trace the birth of their discipline to Graunt's remarkable work. Rothman (1996) proffers the following lessons modern epidemiologists can learn from Graunt:

- He was brief.
- He made his reasoning clear.
- He subjected his theories to repeated and varied tests.
- He invited criticism of his work.
- He was willing to revise his ideas when faced with contradictory evidence.
- He avoided mechanical interpretations of data.

Despite his brilliance with numbers, John Graunt was not a good money manager. He died bankrupt on Easter-eve 1674 and was buried under what was then a pigsty in St. Dunstan's Church in Fleet Street. His eulogy read, ''what pitty 'tis so great an ornament of the city should be buryed so obscurely!'' (Aubrey, 1949).

Germ theory

The notion of a living agent as a cause of disease had been around since ancient times. For instance, the Roman poet Lucretius (circa 100 BC) refers to the seeds

of disease passing from healthy to sick individuals in the poem *De Rerum Natura*. However, the first cogent germ theory was presented by **Girolamo Fracastoro** in 1546 (Saracci, 2001).

Despite early theories of contagion, the prevailing theories of epidemics in the 19th century were expressed in terms of ''spontaneous generation'' and ''miasma atmospheres.'' This manner of thinking began to change midcentury when in 1840 **Jakob Henle** (1809–1885) presented his treatise of the *contagium animatum* in which he theorized that a living substance multiplied within the body where it was excreted by sick individuals and communicated to healthy individuals.

During the same era, **John Snow** (1813–1858) was independently developing similar ideas about contagion, basing his theories on the epidemiologic and pathophysiologic features of cholera. Among Snow's early epidemiologic observations was how cholera spread along the routes of human commerce and war and was propagated from human to human. Among his pathophysiologic observations was the cholera was primarily a gastrointestinal disease and that the loss of fluids caused its systemic effect by means of ''internal congestion'' (sludging of the blood and hypovolemic shock). Snow's theory of contagion recognized that infection with a stabile living organism was necessary for transmission to occur and that the infectious agent multiplies after infections to produce its effects (Winkelstein, 1995). Later in this chapter we will discuss three of Snow's seminal epidemiologic studies.

The French chemist **Louis Pasteur** (1822–1895) ultimately put the doctrine of spontaneous generation to rest by demonstrating that fermentation and organic decay were produced by microorganisms. Pasteur was also the first to isolate an agent responsible for an epidemic disease (in silk worms, in 1865), found that septicemia was caused by anaerobic bacterium, and developed the process for killing germs by heating that still bears his name (''pasteurization'').

Henle's student **Robert Koch** (1843–1910) made a breakthrough when he decided to stain microbes with dye, enabling him to visualize the microbe that caused tuberculosis in 1882 and the cholera bacillus in 1883. Koch is also known for his Postulates, which he developed in 1890.

Until the discovery of arthropod (insect borne) transmission of Texas cattle fever, the only known modes of transmission for infectious agents were by water and air. In 1882, **Daniel E. Salmon** (1850–1914) realized that Texas cattle fever presented something unusual—the disease stayed below a geographic line that extended through the southern United States and Mexico (Figure 1.5) and was not conveyed from bovine to bovine directly or through the atmosphere. Using various epidemiologic and laboratory methods, he and a team of workers at the U.S. Department of Agriculture conducted a series of experiments that demonstrated the vector-borne transmission of the disease. This was the first demonstration of a complex web of causation involving an agent (*Babesia bigeminal*) being transmitted to a mammalian host (cattle) through an invertebrate vector (the tick *Boophilus angulatus*). Discoveries of invertebrate vectors for other diseases (e.g. malaria, yellow fever) soon followed. The complex interactions involved in the maintenance and transmission of an agent in the environment provided the first theories of **medical ecology**.

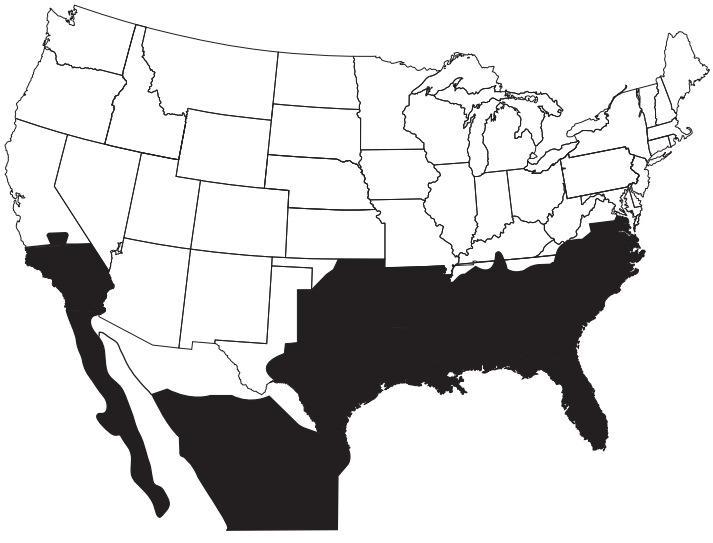


Figure 1.5 Distribution of the *Boophilus* tick before eradication.

Médecine d'observation and La Méthode Numerique (Pinel and Louis)

Owing to a confluence of strong social changes and the consolidation of statistical and probability theory, 18th century France was the incubator of many modern statistical principles and ideas. While the *Academie Royales des Sciences de Paris* were debating Laplace's theory of probability, a parallel movement emphasizing clinical quantification was brewing in the Parisian schools of medicine. The best known of these French physicians were Philippe Pinel and Pierre Charles Alexandre Louis.

Philippe Pinel (1745–1826), primarily known as a pioneer in the scientific and humane treatment of mental illness, also had a passion for medical statistics. Pinel's main statistical achievement was insistence on careful observation and refusal to get lost in undue reliance on unconfirmed theory and appeals to authority. In the introduction to his major work on mental illness published in 1809, he writes that ''a wise man has something better to do than to boast of his cures, namely to be always self-critical.'' After explaining his statistical approach, Pinel states that ''doctors who disapprove of my methods are at liberty to use the method they normally adopt, and a single comparison will suffice to show where the advantage lies'' (Armitage, 1983, p. 322).

In 1795, Pinel was appointed to administer a notorious women's asylum (the Salpêtrière). During his tenure in this position, he collected data on 1002 patients admitted during a 3 year and 9 month period. His studies at the Salpêtrière included cross-classifying cases by year of admission, clinical diagnoses, characteristics of patients at time of admission, and selected outcomes. Using this information, he demonstrated that his overall cure rates were better than those seen in institutions following less enlightened methods. This was true, he concluded, despite the fact that his patient mix tended to have more severe conditions than the comparable institutions. Thus, Pinel was aware of the statistical problem we now call confounding and was able to reason an enlightened approach to its consideration.

Often considered the ''father of clinical statistics,'' the influential French physician **Pierre-Charles Alexandre Louis** (1787–1872; Figure 1.6) wrote: ''I conceive that without the aid of statistics nothing like real medical science is possible.'' Although P.C.A. Louis made careful quantitative observations on many diseases, perhaps his best remembered research evaluated bloodletting as a treatment for various ailments (Louis, 1837).

Bloodletting, an extremely popular form of therapy at the time, required the removal of blood from the patient by lancet or through the placement of leeches on specific parts of the body. The procedure was so popular that 42 million leeches were imported into France in 1833. Louis was the first to call into question the effectiveness of this age-old remedy. Through attentive recordings of clinical observations (*médecine d'observation*), Louis tabulated the response to bloodletting in patients by carefully monitoring the outcome in various treatment groups.

In one analysis, Louis compared death rates and duration of disease in patients who received early treatment (within the first four days of symptoms) and in those

Figure 1.6 Pierre-Charles Alexandre Louis (1787–1872) (Source: Wikipedia Commons).

1	2	3	4	5	6	7	8	9
10 3	7 3	19 3	19 3	28 2	13 1	24 2	19 2	35 1
12 2	10 2	29 3	12 2	17 3	16 2	12 4	12 1	11 2
14 2	12 2	20 2	15 2	40 2	23 3	19 2	18 1	17 2
		20	22 4	13 2	35 5	18 2	20 3	30 3
		16 3	12 4	21 2	17 2	15 2	13 2	
		17 4	21 2	13 2		27 2	21 2	
			25 3					
			28 4					
			40 2					
			16 2					
			12 4					
12 $2\frac{1}{3}$	10 $2\frac{1}{3}$	20 3	20 8	22 2	21 $2\frac{3}{5}$	19 $2\frac{1}{3}$	17 2	23 2

The figures upon the horizontal line above the columns indicate the day when the first bleeding was performed; the figures on the left in each column mark the duration of the diaease; those on the right, the number of bleedings; and those on the horizontal line below, show the mean duration of the disease and the average number of bleedings.

Figure 1.7 Duration of disease and number of bleeding in patients who survived according to day of first treatment. The original legend is reproduced in the figure.

1	2	3	4	5	6	7	8	9
6 5 18	53 5 65	4 1 57	29 2 19	16 4 58	62 4 20	20 2 68	25 1 40	22 1 50
	12 3 69	16 2 54	29 4 46	8 2 63	10 2 40			
	8 2 65	6 3 30	12 1 85	9 4 24	29 3 24			
	12 1 55	6 4 47	15 3 37					
	17 7 75	47 2 75	17 1 67					
		11 4 45	20 3 22					
6 5 18	20 $3\frac{3}{5}$ 66	15 3 51	20 $2\frac{1}{3}$ 49	11 3 48	33 3 28	20 2 68	25 1 40	22 1 50

Figure 1.8 Duration of disease, number of bleeding, and age of patients who died, according to day of first treatment (See Figure 1.7 for meaning of column headings).

who received later treatment (no untreated control group was available). Some of Louis's recordings are shown in Figures 1.7 and 1.8. Using these data, Louis found that mortality was greater in the earlier treated group than in the later treated group (44 versus 25%, respectively).

This type of observation led to the eventual end of this antiquated form of treatment and demonstrated the need for rigorous evaluation of conventional clinical practices. Medical systems that could not withstand a test of observation were to be discredited.

Louis attracted a large following, conveying his beliefs to many of the men who would establish modern medical and public health movements in England, the United States, and continental Europe (Osler, 1897; Lilienfeld and Lilienfeld, 1977). Some of these men were influential in establishing the epidemiologic movement in Victorian England.

The London Epidemiological Society

Urbanization and development of long-distance transportation (shipping and rail) in 19th century Europe led to repeated introductions of cholera, typhoid fever, smallpox, and other infectious diseases into the unsanitary housing conditions of densely populated metropolitan centers. This led to many dramatic outbreaks of ''crowd diseases.'' Driven by pragmatic concerns, a group of English physicians, who realized their obligation went beyond treating sick individuals, chartered the London Epidemiological Society on March 6, 1850 (Lilienfeld, 1978a). The stated objectives in the charter of this organization are remarkably insightful:

> . . . to endeavour, by the light of modern science, to review all those causes which result in the manifestation and spread of epidemic diseases—to discover causes at present unknown, and investigate those which are ill understood—to collect together facts, on which scientific researches may be securely based—to remove errors which impeded their progress—and thus, as far as we are able, having made ourselves thoroughly acquainted with the strongholds of our enemies, and their modes of attack, to suggest those means by which their invasion may either be prevented, or if, in spite of our existence, they may have broken in upon us, to seek how they may be most effectually combated and expelled.
>
> *Babington*, 1850, p. 640

Epidemiology thus came into being as a specific discipline united by its belief that health could be advanced by the scientific study of disease on a population level. One of the members of this new group was a former pupil of P.C.A. Louis: the physician William Farr.

William Farr

It had been nearly two centuries since John Graunt's Observations when, in 1837, the English Parliament created a centralized registration system for information on births, deaths, and marriages. In 1839, William Farr (1807–1883; Figure 1.9) was appointed to head the branch of this office involved with these statistics; he served in this post for the next 40 years. During his tenure, Farr established a national registration system for the collection, classification, analysis, and reporting of mortality statistics—the forerunner of the today's vital statistics and disease surveillance systems. Farr had an insatiable appetite for collecting, tabulating, and analyzing morbidity and mortality statistics. He recognized the importance of standardized nomenclatures of disease, remarking that ''[disease] nomenclature is of as much importance [in epidemiology] as weights and measures in the physical sciences'' (Farr, 1885, p. 234). His anatomically based system of disease classification is the antecedent to the International Classification of Disease currently in use.

Farr relied on comparisons of rates in which numerator data comprised deaths and denominator data comprised population size. Using these simple calculations, Farr compared mortality rates in people of different backgrounds, social classes, and occupations searching for ''causes that make the rates of mortality vary'' (Farr, 1885, p. 187). Figure 1.10 is a replica of one of Farr's tabulations.

Figure 1.9 William Farr (1807–1883).

POPULATION, DEATHS, AND MORTALITY per 1,000 at TWELVE DIFFERENT PERIODS of AGE, in LONDON and in ENGLAND, 1861–70.

AGES.	LONDON. MEAN POPULATION, 1861–1871.		LONDON. ANNUAL DEATHS in 10 Years 1861–70.		LONDON. ANNUAL MORTALITY Per 1,000 living during the Years 1861–70.		ENGLAND. ANNUAL MORTALITY Per 1,000 living during the Years 1861–70.	
	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.
All Ages -	1,415,466	1,613,659	37,581	36,053	26•55	22•34	23•61	21•28
0—	195,963	196,500	17,032	14,997	86•91	76•32	73•16	63•43
5—	161,151	163,821	1,509	1,449	9•37	8•85	8•15	7•76
10—	141,969	145,035	603	590	4•24	4•07	4•46	4•48
15—	131,585	151,530	766	773	5•82	5•10	6•16	6•62
20—	133,185	166,302	1,096	1,034	8•23	6•22	8•45	7•96
25—	233,714	280,674	2,538	2,470	10•86	8•80	9•90	9•69
35—	178,860	206,826	3,066	2,656	17•14	12•84	13•46	12•03
45—	124,417	144,530	3,195	2,676	25•68	18•52	19•17	15•56
55—	71,145	90,739	3,120	3,035	43•85	33•45	33•00	27•77
65—	33,097	49,021	2,741	3,296	82•83	67•23	66•90	58•80
75—	9,325	16,338	1,576	2,394	169•02	146•54	146•58	134•43
85 & upds.	1,055	2,343	339	683	321•42	291•42	313•57	283•64

Figure 1.10 Mortality statistics for London and England in the 19th century. ''The death-rate is a fact; anything beyond this is an inference'' (Source: Farr, 1885, p. 123).

A self-taught mathematician, Farr used actuarial techniques to address questions of mortality and survival. He understood the relation between incidence and prevalence, and was ahead of his time in distinguishing the calculation of risks and rates (Vandenbrouke, 1985). He compared mortality in subgroups to help identify risk factor for morbidity and mortality.

Farr was open-minded about theories of disease etiology. In studying cholera, he initially believed in miasma theory—the false notion that the cholera agent was nonliving and spread through the atmosphere, being ''most fatal at low places.'' However, by 1866, it was clear to Farr that cholera was not transmitted by air but was instead spread by contaminated water (Eyler, 2001).

Farr's theories about the causes of disease included such modern concepts as ''indulgences in excess, by idleness, or by improvidence . . . conflicts with each other . . . organized parasites in the body . . . and molecules which, though of no recognized form, evidently thrive, propagate, die in the bodies of men'' (Farr, 1885, p. 117). The effects of population density on transmission of agents were described, as were properties of herd immunity. Farr also understood the importance of follow-up in evaluating prognosis and the effectiveness of medical treatment (Farr, 1838, 1862). Therefore, it is not surprising to find that Farr has been identified as one of the founders of modern epidemiology (Susser and Adelstein, 1975, p. iii). Farr also provided data and exerted influence on the man who many consider to be the essential hero of modern epidemiology—John Snow.

John Snow

John Snow (1813–1858; Figure 1.11) was a Victorian surgeon with varied scientific and social interests. In addition to being a pioneer in epidemiology, he was a recognized expert in the development and administration of inhaled anesthesia such that he attended the birth of two royal children to administer chloroform gas to Queen Victoria for the purpose of childbirth in 1853 and again in 1857 (Richardson, 1887). Our interests, however, center on his role in epidemiology through his investigations of cholera.

Cholera in Victorian England

Cholera hit Great Britain in 1831–1832, coming from India via the British seaports. As an apprentice to a Newcastle surgeon, Snow attended patients suffering from these early cholera epidemics (Richardson, 1887). When the epidemic resurfaced in 1848, Snow formulated his theories about the disease publishing his views as an article (Snow, 1849a) and booklet (Snow, 1849b). These articles laid out his ideas of cholera as a disease primarily affecting the gastrointestinal tract with the agent entering directly into the alimentary canal orally. Snow theorized that the source of the agent was fecal-contaminated water. This theory contradicted the predominant theory of the time—miasma (''bad air'') theory—which instead professed that cholera arouse from the emanations of inorganic material in the form of foul smelling gases.

Figure 1.11 John Snow (1813—1858). (Source: Wikipedia Commons.)

Miasma theory of transmission

With our current knowledge about infectious diseases transmission, it is difficult to distance ourselves adequately in order to understand the miasma theory of epidemics. In broad strokes however, miasma theory placed emphasis on the noxious vapors produced from ordinary organic decay and decomposition without the presence of prior contagion. This theory was mixed with concepts of ''localizing influences'' which promoted the propagation of the cholera poison, ''predisposing causes,'' ''spontaneous generation,'' and ''cholera atmospheres.'' In contrast to this predominant view of transmission, Snow maintained that ''no mere emanation arising from evolution of foul smelling gases can, *per se*, . . . , originate a specific disease'' (Richardson, 1887, p. xxxix).

Snow's theory

Snow based his theory of cholera pathogenesis on both the clinical and epidemiologic features of the disease. Cholera begins with symptoms specific to the gastrointestinal tract, without the fever and the whole-body signs associated with other epidemic diseases. This caused Snow to postulate ''cholera is, in the first instance at least, a local affection of the mucous membrane of the alimentary canal'' (Snow, 1849a, p. 745). From this, Snow inferred ''the disease must be caused by something which passes from the mucous membrane of the alimentary canal of one patient to that of the other, which it can only do by being swallowed and as the disease grows in a community by what it feeds upon, attacking a few people in a town first, and then becoming more prevalent, it is clear that the cholera poison must multiply itself by a kind of growth'' (Snow, 1849a, p. 746).

Snow also noted that the course of cholera could be traced along with troop movements from India, stating ''one feature immediately strikes the inquirer—viz., the evidence of its communication in human intercourse'' (Snow, 1849a, p. 746). These ideas ultimately coalesced in the form a theory in which Snow proposed that cholera was a self-propagating agent spread from person to person through contaminated water and food.

The London cholera epidemics of 1853–1854 allowed Snow to test these theories using what we now recognize as three distinct epidemiologic methods (Winkelstein, 1995). These are:

- Comparisons of cholera mortality rates using regional aggregate-level data. We now recognize this as the basis of the *ecological study design*.
- Comparison of cholera rates in groups defined by exposure to various water sources. We now recognize this as the basis of the *retrospective cohort study design*.
- Comparison of the characteristics of cholera cases and non-cases in a method that bears some semblance to a case–control study but is really *a case series analysis*.[a]

Snow's ecological analysis

Water distribution in 19th century London was the purview of private water companies. The two major companies that distributed water were the Southwark & Vauxhall Company and the Lambeth Company. During the epidemic of 1849, roughly the same number of deaths occurred in London districts served by either company. However, during the 1853 epidemic, Snow noted that cholera mortality was higher in regions served by the Southwark & Vauxhall Company than in regions served by the Lambeth Company (Figure 1.12), suggesting that water provided by the Southwark & Vauxhall Company served as the vehicle for the dissemination of the cholera agent. This is an ecological comparison because rates are compared by region and there is little or no follow-up of individual experience.

Further investigations led Snow to discover that Southwark & Vauxhall derived its water from downstream sources that were polluted with sewage. In contrast, the Lambeth Company had moved its water source upstream away from the primary sources of sewage pollution, explaining its superior safety.

---

[a] A true case–control study would require a random or at least representative sample of noncases from the population that begat the cases.

Sub-Districts.	Population in 1851.	Deaths from Cholera in 1853.	Deaths by Cholera in each 100,000 living.	Water Supply.
St. Saviour, Southwark	19,709	45	227	Southwark and Vauxhall Water Company only.
St. Olave . . .	8,015	19	237	
St. John. Horsleydown	11,360	7	61	
St. James, Bermondsey	18,899	21	111	
St. Mary Magdalen .	13,934	27	193	
Leather Market .	15,295	23	153	
Rotherhithe* . .	17,805	20	112	
Wandsworth . .	9,611	3	31	
Battersea . . .	10,560	11	104	
Putney . . .	5,280	—	—	
Camberwell . .	17,742	9	50	
Peckham . . .	19,444	7	36	
Christchurch, Southwk	16,022	7	43	Lambeth Water Company, and Southwark and Vauxhall Company.
Kent Road . .	18,126	37	204	
Borough Road . .	15,862	26	163	
London Road . .	17,836	9	50	
Trinity, Newington .	20,922	11	52	
St. Peter, Walworth .	29,861	23	77	
St. Mary, Newington .	14,033	5	35	
Waterloo (1st part) .	14,088	1	7	
Waterloo (2st part) .	18,348	7	38	
Lambeth Church (1st part) . . .	18,409	9	48	
Lambeth Church (2nd part) . . .	26,784	11	41	
Kennington (1st part)	24,261	12	49	
Kennington (2nd part)	18,848	6	31	
Brixton . . .	14,610	2	13	
Clapham	16,290	10	61	
St. George, Chamberwell	15,849	6	37	
Norwood . . .	3,977	—	—	Lambeth Water Company only.
Streatham. . .	9,023	—	—	
Dulwich . . .	1,632	—	—	
First 12 sub-districts .	167,654	192	114	Southwk. & Vaux.
Next 16 sub-districts .	301,149	182	60	Both Companies.
Last 3 sub-districts .	14,632	—	—	Lambeth Comp.

* A part of Rotherhithe was supplied by the Kent Water Company: but there was no cholera in this part.

Figure 1.12 Snow's ecological data on cholera rates by water district, 1853. (Source: Snow, l855, p. 73.)

Snow's retrospective cohort analysis

Snow noticed that there were sub-districts in London where water pipes traveled side-by-side down streets supplying water to households of various sorts. By determining the water supplies for each house and the household of each case, Snow was able to tabulate cholera mortality rates according to water supplier. He found 1263 cholera deaths in the 40 046 households exposed solely to water from the Southwark & Vauxhall (S&V) Company. Thus, households supplied by S&V had a cholera mortality rate of:

$$\frac{1263 \text{ cases}}{40\,046 \text{ households}}$$

$$= 0.0315 \text{ per household or, equivalently, } 315 \text{ per } 10\,000 \text{ household}$$

In households served solely by the Lambeth Company, 98 cholera deaths occurred in 26 107 households, for a rate of:

$$\frac{98 \text{ cases}}{26\,107 \text{ households}} = 0.0038 \text{ or } 38 \text{ per } 10\,000 \text{ households}$$

The rest of London experienced

$$\frac{1422 \text{ cases}}{256\,423 \text{ households}} = 0.0055 \text{ or } 55 \text{ per } 10\,000 \text{ households}$$

Thus, the households supplied by the S&V water company experienced cholera at eight-times the rate of those supplied by the Lambeth water company and more than five-times the rate of the rest of London. This type of analysis of rates according to exposure status forms the basis of the cohort study design.

Snow's case series

In contrast to the cohort method in which rates are compared according to exposure status, case–control studies compare the characteristics of diseased and non-diseased individuals. As part of Snow's inquiry into the terrible outbreak of cholera that affected the Golden Square Area of London in August and September of 1854, he prepared a map showing the distribution of cases in relation to the infamous Broad Street pump (Figure 1.13).

Snow's case series analysis

Snow found that 61 of the fatalities during this outbreak had used water from the Broad Street pump, six had reportedly not drunk water from the pump, and six could not determine whether or not they had used water from the pump. Thus, exposure to Broad Street pump water, if not universal, was very common.

Snow also interviewed cases that seemed to contradict the normal pattern of infection. One interesting observation came when investigating a couple of cases from the suburb of Hampstead, whereby Snow wrote:

> I was informed by this lady's son that she had not been in the neighbourhood of Broad Street for many months. A cart went from Broad Street to West End every day, and it was the custom to take out a large bottle of the water from the pump in Broad Street, as she preferred it. The water was taken on Thursday, 31st August, and she drank of it in the evening, and also on Friday. She was seized with cholera on the evening of the latter day, and died on Saturday, as the above quotation from the register shows. A niece, who was on a visit to this lady, also drank of the water; she returned to her residence, in a high and healthy part of Islington, was attacked with cholera, and died also. There was no cholera at the time, either at West End or in the neighbourhood where the niece died.
>
> *Snow*, 1855; 1936 reprint, pp. 45–46

Thus, although residing outside of the epidemic area, these two cases were discovered to have imbibed water from the Broad Street pump after all.

Snow also interviewed noncases from subpopulations in which cholera had been surprisingly infrequent. For example, in the brewery near the Broad Street pump (see map), no workers had died of cholera. Snow remarked, ''The men are allowed

Figure 1.13 Snow's map of the 1854 Golden Square cholera outbreak. Each horizontal line represents a cholera death. Public water pumps are shown as enclosed dots (⊙). The Broad Street pump is in the center of the map. (Source: Snow, 1855, 1936 reprint, pp. 44 and 45.)

a certain quantity of malt liquor, and Mr. Huggins [the proprietor] believes they do not drink water at all; and he is quite certain that the workmen never obtained water from the pump in the street'' (Snow, 1855, 1936 reprint p. 42).

Publication

As a result of his investigation of the Golden Square (''Broad Street Pump'') outbreak, Snow was able to convince the vestrymen of the parish to remove the handle from the offensive pump. The pump handle was removed, and the plague recessed. More importantly, Snow's lucid observations and reasoning continue to inspire epidemiologists, while his efforts to remove the pump handle serves as a symbol of public health action.

Twentieth-century epidemiology

Many social and scientific events have influenced the development of epidemiology in the 20th century. Industrialization and economic development accelerated greatly, two world wars occurred, the 1918–1919 influenza pandemic claimed between 20 and 40 million lives, European colonialism dissolved, the stock market crashed and a great economic depression ensued, capitalism and communism clashed in a cold war, communism collapsed, world population growth accelerated, medical technology entered into a new stage, communication evolved, networks expanded, life expectancy increased dramatically, and the age structure of populations in industrialized countries transitioned. Concurrent with these trends, epidemiology developed from a descriptive field to an analytic discipline, with biostatistics increasingly serving as an essential discipline (Gordon, 1952).

About mid-century, Wade Hampton Frost, the first professor of epidemiology in the USA, declared that contemporary events had ''extended the meaning of [the word] epidemiology beyond its original limits, to extend not merely the doctrine of epidemics but a science of broader scope in relation to the mass phenomena of disease in their usual or endemic as well as their epidemic occurrence'' (Frost, 1941). Let us consider several early 20th century epidemiologists that led the way in the transition of the discipline.

Emile Durkheim

Emile Durkheim (1858–1917) was a French sociologist known for his compelling scientific approach to studying social phenomena. In his *Rules of Sociological Method* (1895), he sets forth that (a) social explanations require comparisons, (b) comparisons require classification, and (c) classification requires the definition of those facts to be classified, compared, and ultimately explained. Consistent with these rules, Durkheim warned against *notiones vulgares*—the idea that crudely formed concepts of social phenomena without scientific reflection produce only false knowledge: as alchemy had preceded chemistry and astrology had preceded astronomy, untested thoughts on social phenomena merely foreshadows true social science.

Durkheim's seminal work *Le Suicide* (1897) considered many potential risk factors for suicide, including psychopathological states, race, heredity, climate, season, imitative behavior, religion, social instability, and a host of other social phenomena. Table 1.6 is based on Table XXI from *Le Suicide*. From these data, Durkheim concluded: (a) marriage before the age of 20 (''too early marriages'') has an aggravating influence on suicide, especially in men; (b) after age 20, married persons of both sexes enjoy some protection from suicide in comparison with unmarried people; (c) the protective effect of marriage is greater in men; and (d) widowhood diminishes the protective effects of marriage but does not entirely eliminate it. Durkheim reflected on whether the apparent protective effects of marriage were due to the influence of the married domestic environment or whether this ''immunity'' is due to some sort of ''matrimonial selection'' (i.e. people who marry have certain physical and moral constitutions that make them less likely to commit suicide). This type of reflective reasoning and careful interpretation of empirical data foreshadowed the modern epidemiologic approach.

Table 1.6 Rates (per million per year) and relative risks of suicide by age and marital status, France, 1889–1891.

Ages	Rate per million/year			Relative risk	
	Unmarried	Married	Widowed	Unmarried with reference to married	Unmarried with reference to widowed
Men					
15–20	113	500	—	0.22	—
20–25	237	97	142	2.40	1.66
25–30	394	122	412	3.20	0.95
30–40	627	226	560	2.77	1.12
40–50	975	340	721	2.86	1.35
50–60	1434	520	979	2.75	1.46
60–70	1768	635	1166	2.78	1.51
70–80	1983	704	1288	2.81	1.54
Above 80	1571	770	1154	2.04	1.36
Women					
15–20	79.4	33	333	2.39	0.23
20–25	106	53	66	2.00	1.60
25–30	151	68	178	2.22	0.84
30–40	126	82	205	1.53	0.61
40–50	171	106	168	1.61	1.01
50–60	204	151	199	1.35	1.02
60–70	189	158	257	1.19	0.77
70–80	206	209	248	0.98	0.83
Above 80	176	110	240	1.60	0.79

Source: Durkheim, (1897, p. 178, Table XXI).

Joseph Goldberger

Joseph Goldberger (1874–1929; Figure 1.14) was born in the Austrian–Hungarian Empire in a town now located in the Czech Republic. His family emigrated to the United States when he was 6, and settled in Manhattan's Lower East Side. After obtaining his medical degree in 1895 and a brief stint in private practice, he entered the Marine Hospital Service in 1899.[b]

As a young public health officer, Goldberger was assigned to investigate various tropical diseases such as yellow fever, typhoid, and dengue fever, which were the main concerns of the Public Health Service at that time. In 1914, the Surgeon General of the United States appointed Goldberger to investigate the crisis of pellagra which was raging in the southern United States. We now know that pellagra is a nutritional disease caused by severe deficiencies of niacin and the amino acid tryptophan. (The body can synthesize niacin using tryptophan as a precursor.) However, at the time, pellagra was thought to be contagious. Goldberger contradicted this commonly held belief, basing his understanding on the observation that pellagra demonstrated a preference for inmates in hospitals and orphanages, leaving employees of the

[b] The U.S. Marine Hospital Service was established in 1798 to care for seamen and to serve as a bulkhead against infectious agents. It is the forerunner of the Public Health Service, which was established in 1902.

Figure 1.14 Joseph Goldberger (1874–1929).

institutions largely unaffected. Since germs would not distinguish between inmates and employees, Goldberger searched for an alternative cause.

By the spring of 1914, Goldberger had begun his investigations on nutrition and pellagra. Among Goldberger's many studies were nutritional analyses of affected and unaffected households. Table 1.7 is based on a table from Goldberger's 1918 article. This table documents the relative paucity of meats, dairy products, and green vegetables in households with pellagra. Goldberger's work led to nutritional interventions that were effective in treating and preventing pellagra. Note that the bulk of Goldberger's work occurred one to two decades before Elvehjem *et al*. (1937) identified niacin as the specific nutritional deficiency that causes pellagra.

Table 1.7 Caloric intake of foods constituting the average daily supply in nonpellagrous and pellagrous households in seven cotton mill villages, 15-day period in 1916.

Groups of foods	Nonpellagrous households		Pellagrous households	
	Highest income	Lowest income	Lowest income one or more cases	Low-income two or more cases
Meats (exclusive of salt pork), eggs, milk, butter, cheese	762	639	338	270
Dried and canned peas and beans (exclusive of canned string beans)	126	113	115	123
Wheaten flour, bread, cakes and crackers, cornmeal, grits, canned corn, rice	2162	2082	1752	1840
Salt pork, lard, and lard substitutes	741	673	748	745
Green and canned vegetables (exclusive of corn), green and canned string beans, fruits of all kinds	131	71	60	69
Irish and sweet potatoes	55	53	53	46
Sugar, syrup, jellies and jams	250	205	222	217
All foods...	4267	3836	3288	3310

Source: Goldberger *et al*. (1918).

By midcentury, epidemiologic theory and methods took major steps forward in order to study the non-infectious causes of diseases that comprised the major causes of morbidity and mortality as the century progressed. Major advancements were made first in the study of cigarette-related diseases, heart disease, mental disorders, cancers, and medical safety and effectiveness. One of the studies that signaled the reckoning of this new era of ''modern epidemiology'' is the British Doctors Study.

The British Doctors Study

The work of the British team of **Austin Bradford Hill** (1897–1991) and **Richard Doll** (1912–2005) extended many epidemiologic methods in the years following World War II. Bradford Hill's contributions included the introduction of the randomized clinical trial for measuring the benefits of medical interventions, advancements in case–control and cohort methods for the study of exposure–disease relations in observational studies, and articulation of a framework for judging causality using nonexperimental data. Richard Doll's work has been important in transforming our understanding of smoking and other environmental causes of cancer.

As an example, Doll and Hill published an early case–control study linking cigarette smoking to lung cancer in 1950, in which they found a significantly higher proportion of heavy smokers in their case series than in their control series. For instance, 26% of the male lung cancer patients smoked 25 cigarettes a day or more, in comparison to 13% of their control group (Figure 1.15). A similar pattern was found in female cases and controls.

Not long after publishing their case–control study, Doll and Hill sent out inquiries to medical doctors in the United Kingdom asking them to classify their smoking status and quantify the approximate amount they smoked. This brief questionnaire was

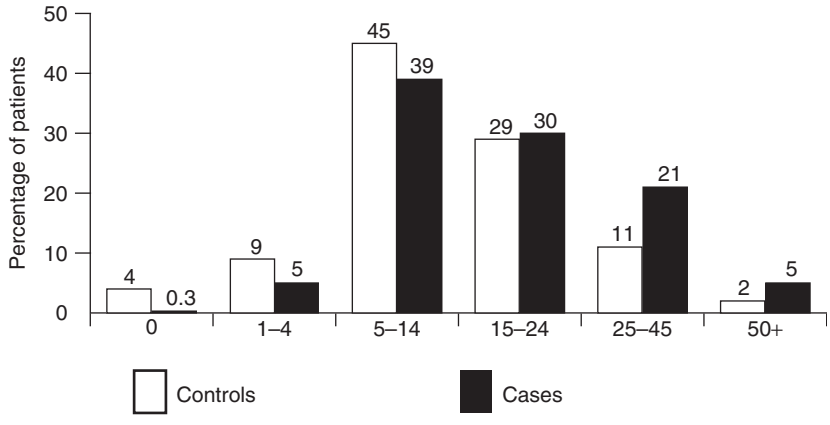


Figure 1.15 Smoking experience of male cases and controls in an early case–control study of lung cancer and smoking (Doll and Hill, 1950).

Table 1.8 Age-adjusted mortality per 1000 person-years according to amount smoked, British Doctors Study.

		Death rates of men smoking a daily average of:			
Cause of death	No. of deaths	Nonsmokers	1–14 g	15–24 g	25+ g
Lung cancer	36	0.00	0.48	0.67	1.14
Other cancers	92	2.32	1.41	1.50	1.91
Respiratory diseases other than cancers	54	0.86	0.88	1.01	0.77
Coronary thrombosis	235	3.89	3.91	4.71	5.15
Other cardiovascular disease	126	2.23	2.07	1.58	2.78
Other diseases	247	4.27	4.67	3.91	4.52
All causes	789	13.61	13.42	13.38	16.30

Source: Doll and Hill (1954).

sent to 59 600 physicians, of which 40 564 replies that were sufficiently complete for analysis were returned. The first report from this cohort study, published in 1954, showed that lung cancer mortality paralleled the amount smoked (Table 1.8). It also showed higher rates of coronary heart disease and other cancers in smokers.

A follow-up report published in 1956 confirmed these smoking-related associations while demonstrating additional associations for chronic obstructive pulmonary disease, peptic ulcer, and pulmonary tuberculosis. After 40 years of follow-up, the British Doctors Study is still ongoing. Figure 1.16 exhibits survival curves for the cohort, demonstrating that 50% of heavy smokers died before age 70 compared with only 20% of nonsmokers; 8% of heavy smokers have survived to age 85, compared with 33% of nonsmokers.

These and other developments following World War II have occurred in the context of rapid growth in understandings of disease etiology. A new age of *modern epidemiology* was thus born, with epidemiology as a discipline distinct from other scientific endeavors with its own concepts and theories, yet still intimately attached to these other disciplines in the study of disease etiology.

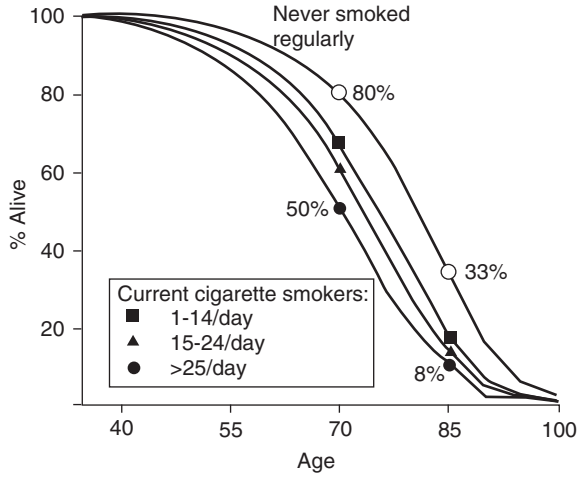


Figure 1.16 Survival in the British Doctors cohort according to amount smoked (Based on data in Doll *et al.*, 1994).

1.4 Chapter summary

Because this is a long chapter with a lot of detail, this summary is provided for review.

Epidemiology and its uses

Epidemiology is the study of the distribution and determinants of health and disease in populations. It is one of the core disciplines of public health, with its objective *to learn* about those factors the prevent disease and injury and promote health. In contrast to clinical medicine, epidemiology focuses primarily on aggregates as opposed to individual patients. Epidemiology is characterized by a close connection between the scientific study of disease causation and application of this knowledge to prevent disease and improve health. It thus covers a broad range of activities, including conducting biomedical research, communicating research findings, and participating with other disciplines in public health interventions. Applications of epidemiology include studying population-based trends in morbidity and mortality, diagnosing health problems in communities, studying the effectiveness of health care, estimating individual chances of disease recovery, identifying new syndromes and characterization of the full spectrum of known ailments, and, most importantly, elucidating the causes of ill-health.

Evolving patterns of morbidity and mortality

Whereas morbidity and mortality in the 19th and early 20th centuries were dominated by acute and infectious causes, the major health problems of today are largely chronic and non-infectious. This shift is known as the *epidemiologic transition*. Accompanying this transition has been a change in population age structure known as the

demographic transition. In 1900, life expectancy at birth in the USA was 47 years. In 2007, life expectancy was almost 78 years. Increases in life expectancy have occurred in all groups, with improvements during the first part of the 20th century mostly during childhood, and improvement during the second half of the century largely during middle- and late-age. Concomitant with these changes has been a decrease in birthrates, resulting in an aging of the population.

Selected historical figures and events

Epidemiological insights into health and disease are probably as old as civilization itself. The scientific roots of epidemiology can be traced to Hippocratic principles developed in the 4th century BCE. However, the central tenants of modern epidemiology can be attributed to the renaissance of scientific and artistic ideas starting in the 16th century and developed in the Age of Enlightenment starting in the 18th century. Epidemiology emerged as a unique discipline in Victorian England with the establishment of the London Epidemiological Society in 1850, with the work of many individuals, notably William Farr and John Snow. In the 19th century and first half of the 20th century, epidemiology was concerned primarily with the control of infectious diseases. Beginning in the early 20th century, as the burden of disease shifted from acute infectious diseases to chronic ''lifestyle diseases,'' developments in the epidemiologic study of chronic diseases and medical safety and effectiveness took on greater importance. Rapid growth in understanding the epidemiologic study all types of diseases encouraged a modern form of epidemiology, with its own distinct scientific practices and theories.

Review questions

R.1.1 The word epidemiology is based on the Greek terms *epi*, *demos*, and *ology*. Define each of these terms.

R.1.2 Select your favorite definition of epidemiology. What appeals to you about this particular definition?

R.1.3 How does *epidemiology* differ from *clinical medicine*? How does it differ from *public health*?

R.1.4 The preamble to the 1948 constitution of the World Health Organization addresses three elements of health and well-being. Name these three elements.

R.1.5 Define these terms: *epidemic*, *pandemic*, *endemic*, *morbidity*, *mortality*.

R.1.6 Is it the responsibility of the epidemiologist to effectively communicate their findings? Explain.

R.1.7 List Morris's (1957) seven general uses of epidemiology.

R.1.8 One of Morris's (1957) uses of epidemiology is *community diagnosis*. What does this mean in plain terms?

R.1.9 Describe the demographic transition of the 20th century.

R.1.10 Describe the epidemiologic transition of the 20th century.

R.1.11 Age-adjusted mortality decreased by approximately 40% from 1950 to 2000. Which major causes of death demonstrated the steepest declines?

R.1.12 Describe the change in the shape of the population pyramid that occurred during the 20th century.

R.1.13 List several examples of modifiable risk factors for chronic diseases.

R.1.14 True or false? Overall age-adjusted cancer mortality rates increased dramatically during the second half of the 20th century. Explain your response.

R.1.15 True or false? Age-adjusted cardiovascular mortality rates continued to increase dramatically during the second half of the 20th century. Explain your response.

R.1.16 List the three current most popular causes of death in the USA in rank order.

R.1.17 True or false? Life expectancy at birth increased by about 30 years during the 20th century.

R.1.18 In what century was epidemiology first recognized as a unique discipline?

R.1.19 Which ancient philosopher/physician is said to have initially freed the study of health and disease from philosophical speculation, superstition, and religion?

R.1.20 List features of scientific work.

R.1.21 Match each of these historical figures with their brief biographical descriptions.
Historical figures: William Farr, Fracastoro, John Graunt, Pierre-Charles Alexandre Louis, Philippe Pinel, Percival Pott, Daniel Salmon, John Snow, Thomas Syndenham
Brief descriptions:

(A) The ''English Hippocrates,'' in the 1600s.
(B) Identified soot as the cause of scrotal cancer in 18th century chimney sweeps.
(C) 17th century physician who used population-based vital statistics to derive early epidemiologic observations.
(D) Presented first cogent germ theory of disease in 1545.
(E) 19th century American veterinarian who led the team that discovered the first vector-borne ecology of a disease.
(F) 18th/19th century French physician who pioneered the humane treatment of mental illness and said ''a wise man has something better to do than to boast of his cures, namely to be always self-critical.''
(G) 19th century French physician whose studies led him to believe that bloodletting was infective in the treatment of pneumonia; known as a proponent of ''the numerical method.''
(H) 19th century British physician who was the first registrar of mortality statistics nationally; pioneer in the use of vital statistics and epidemiologic methods in Victorian England.
(I) Victorian surgeon who studied the transmission of cholera; best known for convincing authorities to remove the handle from the Broad Street pump.

References

Armitage, P. (1983) Trials and errors: The emergence of clinical statistics. *Journal of the Royal Statistical Society. Series A (General)*, **146**, 321–334.

Aubrey, J. (1949) *Aubrey's Brief Lives*. Available: *http://www.york.ac.uk/depts/maths/histstat/aubrey.pdf* (accessed 23 March 2003).

Babington, B. G. (1850) Epidemiological Society. Monday, Dec. 2, 1850. Dr. Babington, President. *Lancet*, **2**, 639–642.

Brownson, R. C., Remington, P. L., and Davis, J. R. (1993) *Chronic Disease Epidemiology and Control*, American Public Health Association, Washington, D.C.

Bureau of the Census (1904) *Abstract of the Twelfth Census of the United States 1900*, 3rd edn, Government Printing Office, Washington, D.C.

CDC/NCHS (2006) Health, United States, 2006: chartbook on trends in the health of Americans. Available from *www.cdc.gov/nchs/data/hus/hus06.pdf*.

CDC/NCHS (2010) Health, United States, 2009: with special feature on medical technology. Available: *www.cdc.gov/nchs/data/hus/hus09.pdf*.

Calhoun, J. G., Ramiah, K., Weist, E. M., and Shortell, S. M. (2008) Development of a core competency model for the master of public health degree. *American Journal of Public Health*, **98** (9), 1598–1607.

Cole, P. (1979) The evolving case–control study. *Journal of Chronic Diseases*, **32**, 15–27.

Doll, R., and Hill, A. B. (1950) Smoking and carcinoma of the lung. *British Medical Journal*, **2**, 739–748.

Doll, R., and Hill, A. B. (1954) The mortality of doctors in relation to their smoking habits. *British Medical Journal*, **4877**, 1451–1455.

Doll, R., and Hill, A. B. (1956) Lung cancer and other causes of death in relation to smoking: a second report on the mortality of British doctors. *British Medical Journal*, ***ii***, 1071–1076.

Doll, R. (1992) Health and the environment in the 1990s. *American Journal of Public Health*, **82**, 933–941.

Doll, R., Peto, R., Wheatley, K., Gray, R., and Sutherland, I. (1994) Mortality in relation to smoking: 40 years' observations on male British doctors. *British Medical Journal*, **309** (6959), 901–911.

Durant, A., and Durant, W. (1961) *The Story of Civilization, 7, the Age of Reason Begins*, Simon & Shuster, New York.

Durkheim, E. (1895) *The Rules of the Sociological Method (I982 Reprint)* (ed. S. Lukes; transl. W. D. Halls), Free Press, New York.

Durkheim, E. (1897) *Suicide: A Study in Sociology (I951 Translation)* (transl. J. A. Spaulding and G. Simpson), Free Press, New York.

Elvehjem, C. A., Madden, R. J., Strong, F. M., and Woolley, D. W. (1937; 1974 reprint) The isolation and identification of the anti-black tongue factor 1937. *Nutrition Reviews*, **32** (2), 48–50.

Evans, A. S. (1979) Re: ''Definitions of Epidemiology.'' *American Journal of Epidemiology*, **109**, 379–382.

Eyler, J. M. (2001) The changing assessments of John Snow's and William Farr's cholera studies. *Sozialund Präventivmedizin*, **46**, 225–232.

Farr, W. (1838) On prognosis. *British Medical Almanac, Suppl*., 199–216.

Farr, W. (1862) A method of determining the effects of systems of treatment in certain diseases. *British Medical Journal*, **2**, 193–195.

Farr, W. (1885) *Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr (1975 Reprint)*, Offices of the Sanitary Institute, Metuchen, NJ.

Frost, W. H. (1941) Risk of persons in familial contact with pulmonary tuberculosis (Reprint of *Am. J. Pub. Health*, 1933; **23**:426–432), in *Papers of Wade Hampton Frost, M.D.*: *A Contribution to Epidemiological Method* (ed. K. F. Maxcy), The Commonwealth Fund, New York, pp. 582–591.

Gordon, J. E. (1952) The twentieth century—yesterday, today and tomorrow (1920–), in *The History of American Epidemiology* (ed. C.-E. A. Winslow), CV Mosby; St. Louis, pp. 114–167.

Graunt, J. (1662) *Natural and Political Observations Mentioned in a Following Index and Made Upon the Bills of Mortality* (web reprint), *http://www.ac.wwu.edu/~stephan/Graunt/bills.html*.

Greenwood, M. (1935) *Epidemics and Crowd-Diseases: An Introduction to the Study of Epidemiology*, Williams & Norgate, London.

Hippocrates. (400 BCE) On airs, waters, and places, in *The Internet Classics Archive* (transl. F. Adams), retrieved 21 June 2012, from *http://classics.mit.edu/Hippocrates/airwatpl.html*.

Institute of Medicine (1988) The future of public health. Paper presented at the Committee for the Study of the Future of Public Health, Washington, D.C.

Kuller, L. H. (1991) Epidemiology is the study of ''epidemics'' and their prevention. *American Journal of Epidemiology*, **134**, 1051–1056.

Last, J. M. (ed.) (2001) *A Dictionary of Epidemiology* (4th edn), Oxford University Press, Oxford.

Lilienfeld, D. E., and Lilienfeld, A. M. (1977) Epidemiology: A retrospective study. *American Journal of Epidemiology*, **106**, 445–459.

Lilienfeld, D. E. (1978a) ''The greening of epidemiology'': Sanitary physicians and the London Epidemiological Society (1830–1870). *Bulletin of the History of Medicine*, **52**, 503–528.

Lilienfeld, D. E. (1978b) Definitions of epidemiology. *American Journal of Epidemiology*, **107**, 87–90.

Louis, P. C. A. (1837) *Researches on the Effects of Bloodletting in Some Inflammatory Diseases* (transl. C. G. Putnam), Hilliard, Gray, and Co., Boston.

Mausner, J. S., and Baum, A. K. (1974) *Epidemiology*, W.B. Saunders, Philadelphia.

Miettinen, O. S., and Flegel, K. M. (2003) Elementary concepts of medicine: V. Disease: one of the main subtypes of illness. *Journal of Evaluation in Clinical Practice*, **9** (3), 321–323.

Morris, J. N. (1957) *Uses of Epidemiology*, Livingstone, Edinburgh.

National Office of Vital Statistics (1947) *Deaths and Death Rates for Leading Causes of Death: Death Registration States, 1900–1940*. Photocopy received from National Center for Health Statistics, Hyattsville, MD (8 March 1996).

O'Donnell, T. (1936) *History of Life Insurance in its Formative Years; Compiled from Approved Sources*, American Conservation Company, Chicago.

Omran, A. R. (1971) The epidemiologic transition: A theory of the epidemiology of population change. *The Milbank Memorial Fund Quarterly*, **49**, 509–538.

Osler, W. (1897) Influence of Louis on American Medicine. *Bulletin of the Johns Hopkins Hospital*, **VIII**, 161–167.

Porta, M. S. (ed.) (2008) *A Dictionary of Epidemiology*, 5th edn, Oxford University Press, Oxford.

Pott, P. (1790) *The Chirurgical Works of Purcivall Pott*, Hanes W. Clark and R. Collins, London (original work published 1775).

Ramazzini, B. (1713; 2001) De morbis artificum diatriba [Diseases of workers]. Sections published in *American Journal of Public Health*, **91** (9), 1380–1382.

Richardson, B. W. (1887) John Snow, M.D. A representative of medical science and art of the Victorian era. (Reprinted in the 1936 text *Snow on cholera*). *The Asclepiad*, **IV**, 274–300.

Salmon, D. E. (1882) The dissemination of Texas fever, and how to control it. U.S.D.A Special Report 50.

Saracci, R. (2001) Introducing the history of epidemiology, in *Teaching Epidemiology—A Guide for Teachers in Epidemiology, Public Health and Clinical Medicine* (eds J. Olsen, R. Saracci, and D. Trichopoulos), Oxford University Press, New York, pp. 384.

Snow, J. (1849a) On the pathology and mode of communication of cholera. *London Medical Gazette*, **44**, 745–752.

Snow, J. (1849b) *On the Mode of Communication of Cholera*. John Churchill, London.

Snow, J. (1855) *On the Mode of Communication of Cholera*, 2nd edn, much enlarged, John Churchill, London.

Snow, J. (1855; reprint) *Snow on Cholera*, being a reprint of two papers by John Snow, M.D., together with a biographical memoir by B. W. Richardson, M.D., and an introduction by Wade Hampton Frost, M.D. (1936), The Commonwealth Fund, New York.

Susser, M. (1973) *Causal Thinking in the Health Sciences. Concepts and Strategies in Epidemiology*, Oxford University Press, New York.

Susser, M., and Adelstein, A. (1975) *Introduction, Vital statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr (1885)*, Scarecrow Press, Metuchen, NJ, pp. iii–xiv.

Vandenbrouke, J.P. (1985) On the rediscovery of a distinction. *American Journal of Epidemiology*, **121** (5), 627–628.

Whitman, W. (1954) *The Complete Poetry and Prose of Walt Whitman*, vol. II, Garden City Books, Garden City, NY.

Winkelstein, W., Jr., (1995) A new perspective on John Snow's communicable disease theory. *American Journal of Epidemiology*, **142** (9 Suppl), S3–S9.

Xu J. Q, Kochanek K. D, Murphy S. L, Tejada-Vera B. (2010.) Deaths: Final data for 2007. *National Vital Statistics Reports*, vol. 58, No. 19, National Center for Health Statistics, Hyattsville, MD.

CHAPTER 2
Causal Concepts

2.1 Natural history of disease
- Stages of disease
- Stages of prevention

2.2 Variability in the expression of disease
- Spectrum of disease
- The epidemiologic iceberg

2.3 Causal models
- Definition of cause
- Component cause model (causal pies)
- Causal web
- Agent, host, and environment

2.4 Causal inference
- Types of decisions
- Report of the Advisory Committee to the U.S. Surgeon General, 1964
- Hill's framework for causal inference
- Philosophical considerations

Exercises

Review questions

References

2.1 Natural history of disease

Stages of disease

The **natural history of disease** refers to the progression of a disease in an individual over time. This includes all relevant phenomena from before initiation of the disease (the **stage of susceptibility**) until its resolution (Figure 2.1). In the period following exposure to the causal factor, the individual enters a **stage of subclinical disease** (also called the **preclinical phase**). For infectious agents, this corresponds to the **incubation period** during which the agent multiplies within the body but has not yet produced discernible signs or symptoms. For noninfectious diseases, this corresponds to the **induction period** between a causal action and disease initiation.

The **stage of clinical disease** begins with a patient's first symptoms and ends with resolution of the disease. Be aware that the onset of symptoms marks the beginning of this stage, not the time of diagnosis. The time-lag between the onset of symptoms and diagnosis of disease can be considerable. **Resolution** of the disease may come by

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

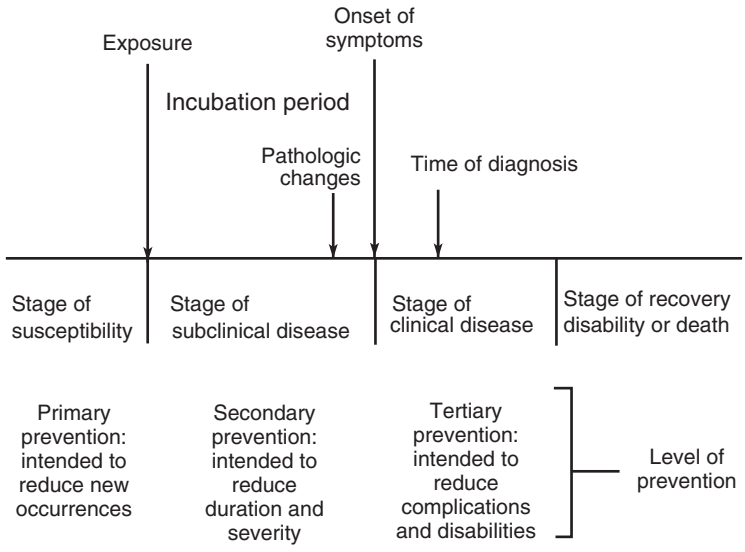


Figure 2.1 Stages in the natural history of disease and levels of prevention.

means of recovery or death. When recovery is incomplete the individual may be left with a disability.

Incubation periods of infectious diseases vary considerably. Some infectious diseases are characterized by short incubation periods (e.g., cholera has a brief 24- to 48-hour incubation period). Others are characterized by intermediate incubation periods (e.g., chickenpox has a typical incubation period of 2–3 weeks). Still others are characterized by extended incubation periods (e.g., the median incubation period of acquired immunodeficiency syndrome (AIDS) can be measured in decades). Table 2.1 lists incubation periods for selected infectious diseases. Note that even for a given infectious disease, the incubation period may vary considerably. For example, the incubation period for human immunodeficiency virus (HIV) and AIDS ranges from 3 to more than 20 years.

Induction periods for noninfectious diseases also exhibit a range. For example, the induction period for leukemia following exposure to fallout from the atomic bomb blast in Hiroshima ranged from 2 to more than 12 years (Cobb *et al.*, 1959). As another example, Figure 2.2 illustrates the empirical induction periods for bladder tumors in industrial dyestuff workers (Case *et al.*, 1954). Variability in incubation is due to differences in host resistance, pathogenicity of the agent, the exposure dose, and the prevalence and availability of cofactors responsible for disease.

Understanding the natural history of a disease is essential when studying its epidemiology. For example, the epidemiology of HIV/AIDS can only be understood after identifying its multifarious stages (Figure 2.3). Exposure to HIV is followed by an acute response that may be accompanied by unrecognized flu-like symptoms. During this acute viremic phase, prospective cases do not exhibit detectable antibodies in their serum, yet may still transmit the agent. During a lengthy induction, CD4+ lymphocyte counts decline while the patient is still free from symptoms. The risk

Table 2.1 Incubation periods for selected infectious diseases.

Disease	Typical incubation period
Acquired immune deficiency syndrome	Infection to appearance of antibodies: 1–3 months; median time to diagnosis: approx. 10 years; treatment lengthens the incubation period
Amebiasis	2–4 weeks
Chickenpox	13–17 days
Common cold	2 days
Hepatitis B	60–90 days
Influenza	1–5 days
Legionellosis	5–6 days
Malaria (*Plasmodium vivax* and *P. ovale*)	14 days
Malaria (*P. malariae*)	30 days
Malaria (*P. falciparum*)	12 days
Measles	7–18 days
Mumps	12–25 days
Poliomyelitis, acute paralytic	7–14 days
Plague	2–6 days
Rabies	2–8 weeks (depends on severity of wound)
Salmonellosis	12–36 hours
Schistosomiasis	2–6 weeks
Staphylococcal food poisoning	2–4 hours
Tetanus	3–21 days

Source: Benensen (1990).

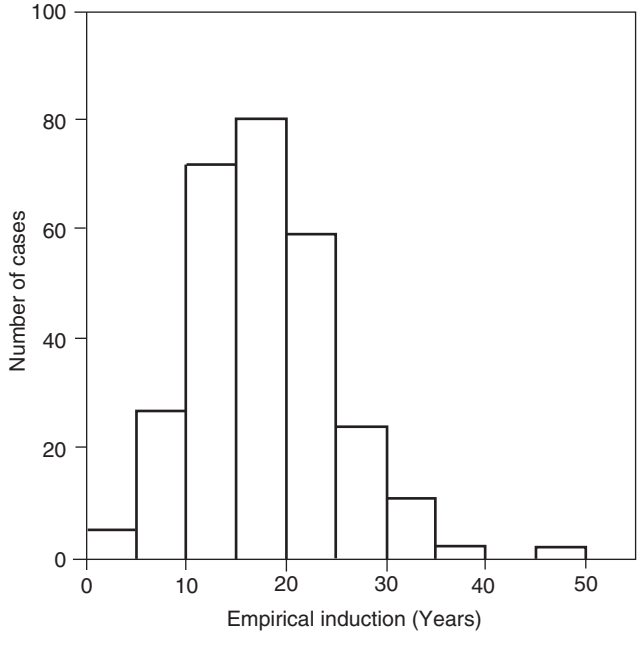


Figure 2.2 Number of years after starting work and onset of urinary bladder tumors in industrial dyestuff workers (Source: Case *et al.,* 1954).

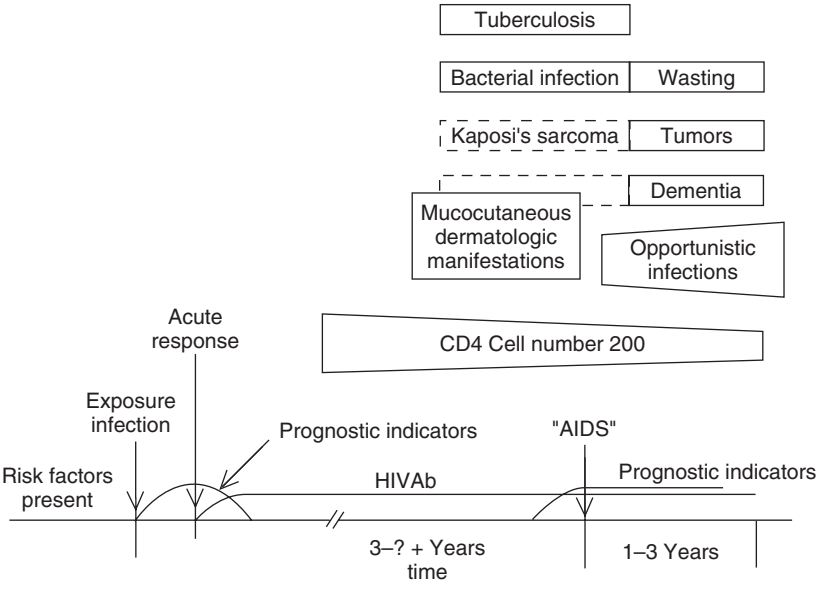


Figure 2.3 Natural history and progression of HIV/AIDS (Source: Cotton, 1995).

of developing AIDS is low during these initial years, but increases over time as the immune response is progressively destroyed, after which AIDS then may express itself in different forms (e.g., opportunistic infections, encephalitis, Kaposi's sarcoma, dementia, wasting syndrome).

A slightly more sophisticated view of the natural history of disease divides the subclinical stage of disease into an induction period and a latent period (Figure 2.4). **Induction** occurs in the interval between a causal action and the point at which the occurrence of the disease becomes inevitable. A **latent period** follows after the disease becomes inevitable but before clinical signs arise. During this latent phase, various causal factors may promote or retard the progression of disease. The induction and promotion stages combined are referred to as the **empirical induction period** (Rothman, 1981). This more sophisticated view better suits the consideration of multi-factored disease, where multiple factors must act together to result in a cause.

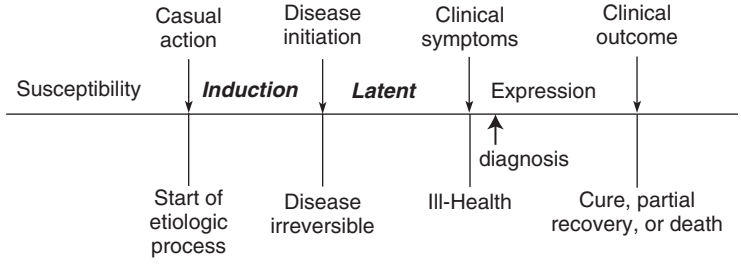


Figure 2.4 Induction period, latent period, and empirical induction period.

Stages of prevention

Disease prevention efforts are classified according to the stage of disease at which they occur (Figure 2.1). **Primary prevention** is directed toward the stage of susceptibility. The goal of primary prevention is to prevent the disease from occurring in the first place. Examples of primary prevention include needle-exchange programs to prevent the spread of HIV, vaccination programs, and smoking prevention programs.

Secondary prevention is directed toward the subclinical stage of disease, after which the individual is exposed to the causal factor. The goal of secondary prevention is to prevent the disease from emerging or delay its emergence by extending the induction period. It also aims to reduce the severity of the disease once it emerges. Treating asymptomatic HIV-positive patients with antiretroviral agents to delay the onset of AIDS is a form of secondary prevention.

Tertiary prevention is directed toward the clinical stage of disease. The aim of tertiary prevention is to prevent or minimize the progression of the disease or its sequelae. For example, screening and treating diabetics for diabetic retinopathy to avert progression to blindness is a form of tertiary prevention.

2.2 Variability in the expression of disease

Spectrum of disease

Diseases often display a broad range of manifestations and severities. This is referred to as the **spectrum of disease**. Both infectious and noninfectious diseases exhibit spectrums. When considering infectious diseases, there is a **gradient of infection**. As an example, HIV infection ranges from inapparent, to mild (e.g., AIDS-related complex), to severe (e.g., wasting syndrome). As an example of a noninfectious disease's spectrum, consider that coronary artery disease exists in as asymptomatic form (atherosclerosis), transient myocardial ischemia, and myocardial infarctions of various severities.

The epidemiologic iceberg

The bulk of a health problem in a population may be hidden from view. This phenomenon, referred to as the **epidemiologic iceberg** (Last, 1963), applies to infectious, noninfectious, acute, and chronic diseases alike.

Uncovering disease that might otherwise be ''below sea level'' by screening and better detection often allows for better control of health problems. Consider that for every successful suicide attempt there are dozens of unsuccessful attempts and a still larger number of people with depressive illness that might be severe enough to have them wish to end their lives. With appropriate treatment, individuals with suicidal tendencies would be less likely to have suicidal ideation and be less likely to attempt suicide. As another example: reported cases of AIDS represent only the tip of HIV infections. With proper antiretroviral therapy, clinical illness may be delayed and transmission averted.

Dog bite injuries provide another example. In 1992 and 1994, there were 20 deaths due to dog bites annually. However, by relying solely on death certificate information,

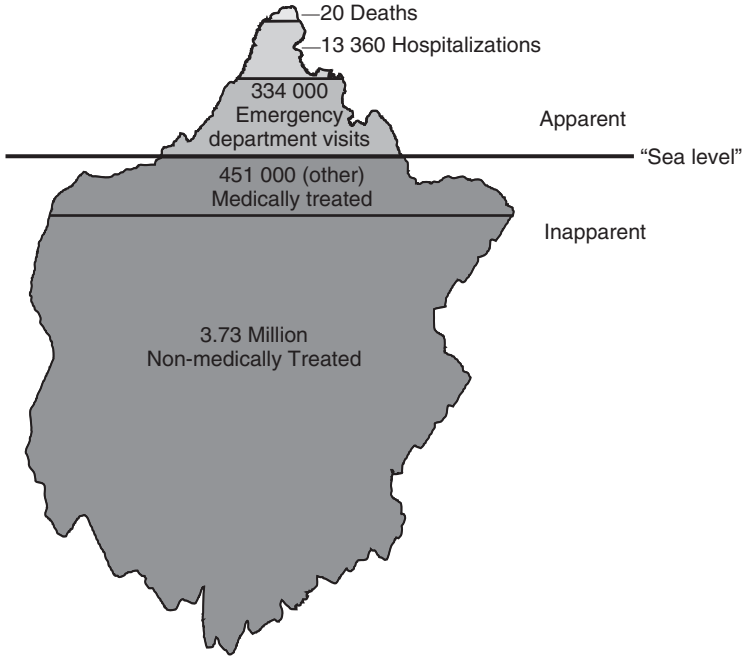


Figure 2.5 Epidemiologic iceberg: annual number of dog bite injuries in the United States, 1992–1994 (Based on Weiss *et al*., 1998).

many additional serious dog bite injuries go undetected. For each fatal dog bite there were 670 dog bite hospitalizations, 16 000 emergency department visits for dog bites, 21 000 medical visits to other clinics, and 187 000 non-treated bites (Weiss *et al*., 1998; Figure 2.5). With recognition of this problem, more effective animal control and surveillance programs can be put into place to prevent future dog bite injuries.

2.3 Causal models

Definition of cause

Effective disease control and prevention depends on understanding the causes of illnesses. In general terms, a **cause** is something that produces an effect or brings about a result. At a deeper level, a cause is

> . . . an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed.
>
> *Hume*, 1772, Section VII

This statement has two essential elements. Firstly, the cause must precede its effect. Secondly, the effect would not have occurred if the cause did not precede it. The causal argument goes something like this: ''if the person who developed disease *Y* had

not been exposed to factor *X*, then disease *Y* would not have occurred. Therefore, *X* is a cause.''

In addition, the modern definition of cause incorporates an important element of time:

> A cause of a disease event is an event, condition or characteristic that preceded a disease without which the disease event either would not have occurred at all or would not have occurred until some later time.
>
> *Rothman and Greenland*, 1998, p. 8

On a population basis, we expect that an increase in the level of a causal factor in inhabitants will be accompanied by an increase in the incidence of disease in that population, *caeteris parabus* (all other things being equal). We also expect that if the causal factor can be eliminated or diminished, the frequency of disease or its severity will decline.

Component cause model (causal pies)

Most diseases are caused by the cumulative effect of multiple causal components acting (''interacting'') together. Thus, a **causal interaction** occurs when two or more causal factors act together to bring about an effect. Causal interactions apply to both infectious and noninfectious diseases and explains, for example, why two people exposed to the same cold virus will not necessarily experience the same outcome: one person may develop a cold while the other person may experience no ill effects.

Rothman's (1976) causal pies helps clarify the contribution of causal components in disease etiology. Figure 2.6 displays two causal mechanisms for a disease. Let us assume these are the only two mechanisms that cause this ailment. Wedges of each pie represent components of each causal mechanism, corresponding to risk factors we hope to identify. Each pie represents a **sufficient causal mechanism**, defined as a set of factors that in combination makes disease occurrence inevitable. Each casual component (wedge) plays an essential role in a given causal mechanism (pie); a specific disease may result from a number of different causal combination mechanisms.

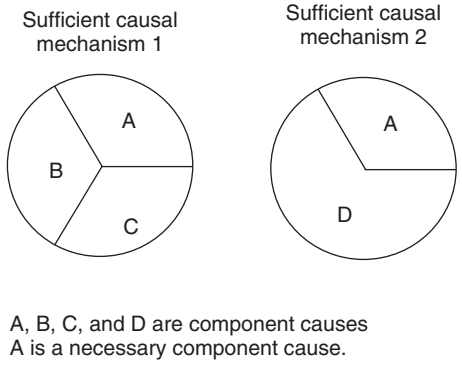


Figure 2.6 Two sufficient causal mechanisms (''pies'').

A causal factor is said to be **necessary** when it is a component cause member of every sufficient mechanism. In other words, the component cause is necessary if the disease cannot occur in its absence. In Figure 2.6, Component *A* is a necessary cause, since it is evident in all possible mechanisms—the disease cannot occur in its absence. For example, the tubercular bacillus *Mycobacterium tuberculosis* is a necessary cause of tuberculosis. However, it is not sufficient by itself to cause disease: it is common for a person to harbor the *Mycobacterium* in their body while remaining disease-free. Some individuals are not susceptible to tuberculosis; they are resistant. Therefore, there are complementary factors that enable disease to manifest. Examples of complementary factors for the manifestation of tuberculosis include familial exposure, immunosuppression, genetic susceptibility, poor nutrition, overcrowding, and high environmental loads of the agent.

Causal components that do not occur in every sufficient mechanism yet are still essential in some cases are said to be **contributing component causes**. For example, cigarette smoking is a contributing but not a necessary cause of lung cancer, since it contributes to the cause of the (vast majority) lung cancer, but is not necessary in every case. (Approximately 5–10% of lung cancer cases occur in nonsmokers.) Likewise, high serum cholesterol, while neither necessary nor sufficient as a cause of coronary heart disease, is an indispensable component of many such causal processes. In Figure 2.6, *B*, *C*, and *D* are nonnecessary contributing causal components.

Component causes that complete a given causal mechanism (pie) are said to be **causal complements**. In Figure 2.6, for example, the causal complements of factor *A* in Mechanism 1 is $(B + C)$. In mechanism 2, the causal complement of factor *A* is *D*. Factors that work together to form sufficient causal mechanism are said to **interact causally**.[a]

Causal interactions have direct health relevance. For example, when a person develops an infectious disease, the causal agent must interact with the causal complement known as ''susceptibility'' to cause the disease. When considering hip fractures in elderly patients, the necessary element of trauma interacts with the causal complement of osteoporosis to cause the hip fracture. In similar veins, smoking interacts with genetic susceptibility and other environmental factors in causing lung cancer, and dietary excesses interact with lack of exercise, genetic susceptibility, atherosclerosis and various clotting factors to cause heart attacks. Causal factors rarely act alone.

Causal pies demonstrate that **individual risk** is an all-or-none phenomenon. In a given individual, either a causal mechanism is or is not completed. This makes it impossible to directly estimate individual risk. In contrast, the notion of **average risk** is a different matter. Average risk can be estimated directly as the proportion of individuals regarded as a member of a recognizable group that develops a particular condition. For example, if one in ten smokers develop lung cancer over their lifetime, we can say that this population has a lifetime risk for this outcome of one in ten.

The effects of a given cause in a population depend on the **prevalence of causal complements** in that population. The effect of phenylketanines, for instance, depends not only on the prevalence of an inborn error of metabolism marked by the absence

[a] The concept of a *causal* interaction is not to be confused with that of a *statistical* interaction, despite the similarity of these terms.

of phenylalanine hydroxylase, but depends also on the environmental prevalence of foods high in phenylalanine. Similarly, the effects of falls in the elderly depend not only on the opportunity for falling, but also on the prevalence of osteoporosis. The population-wide effects of a pathological factor cannot be predicted without knowledge of the prevalence of its causal complements in the population.

Hogben's (1933) example of **yellow shank disease in chickens** provides a memorable example of how population effects of a given causal agent cannot be separated from the prevalence of its causal complements. The trait of yellow shank in poultry is a condition expressed only in certain genetic strains of fowl when fed yellow corn. A farmer with a susceptible flock who switches from white corn to yellow corn will perceive the disease to be caused by yellow corn. A farmer who feeds only yellow corn to a flock with multiple strains of chickens, some of which are susceptible to the yellow shank condition, will perceive the condition to be caused by genetics. In fact, the effects of yellow corn cannot be separated from the genetic makeup of the flock, and the effect of the genetic makeup of the flock cannot be separated from the presence of yellow corn in the environment. To ask whether yellow shank disease is environmental or genetic is like asking whether the sound of a faraway drum is caused by the drum or the drummer—one does not act without the other. This is what we mean by *causal interaction*.

Causal web

The **causal web** is a metaphor that emphasizes the interconnectedness of direct and indirect cause of disease and ill-health. **Direct causes** are proximal to the pathogenic mechanism. **Indirect causes** are distal or ''upstream'' from the disease causing mechanism. Figure 2.7 depicts the well-established causal web for myocardial infarction (heart attack). The direct cause (pathogenic mechanism) of myocardial infarction is coronary artery blockage and subsequent death of the heart muscle. However, this disease also has indirect factors upstream from this direct cause when one considers the social and environmental factors that lead to hyperlipidemia, obesity, a sedentary lifestyle, arteriosclerosis, coronary stenosis, and ultimately to the coronary artery blockage.

Levels of cause in a causal web may broadly be classified as:

- **Macro-level** (indirect causes, such as social, economic, cultural, and evolutionary determinants)
- **Individual-level** (intermediate-level cause, such as personal, behavioral, and physiological determinants)
- **Micro-level** (direct cause at the organ, cellular, and molecular level).

Consider, for example, the cause of early childhood mortality in non-industrialized countries. In this example, macro-level causes encompass broad social, economic, and cultural conditions that lead to a paucity of clean water, food, shelter, and sanitation. Individual-level causes include child-care practices that expose children to pathogens, malnutrition, and dehydration. Micro-level causes include the immediate pathophysiologic interaction between malnutrition and the pathogenic respiratory and gastrointestinal agents that ultimately lead to death (Millard, 1994).

The relative contribution of these various levels of cause in epidemiology and public health have been the subject of considerable and sometimes contentious debate, with

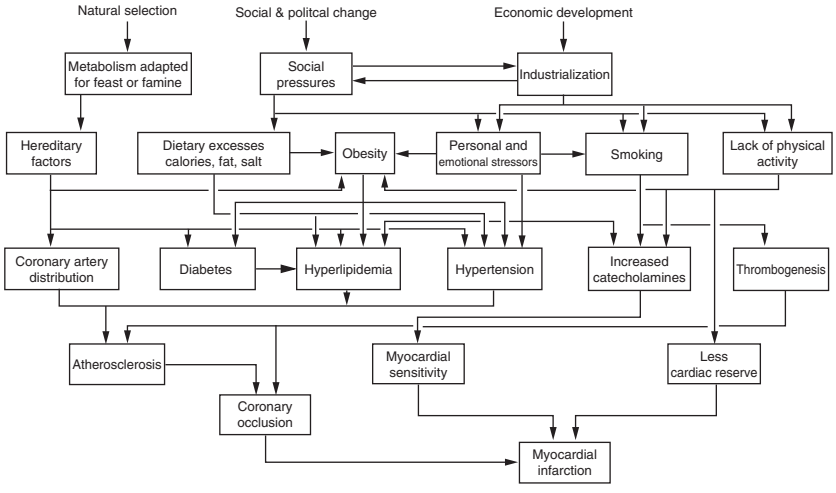


Figure 2.7 Causal-web model for myocardial infarction.

advocates for each level claiming particular and profound benefits for their way of addressing problems. In practice, however, advocating one or another level may hinder achieving the most practical solution for preventing a given public health problem. Maintaining fragmented methods of research into the various levels of cause can only obstruct our understanding and ultimately delay effective prevention measures (Savitz, 1997).

Agent, host, and environment

Causal components can be classified as agent, host, or environmental factors (Figure 2.8). **Agents** are biological, physical, and chemical factors whose presence, absence, or relative amount (too much or too little) are necessary for disease to occur (Table 2.2). **Host factors** include personal characteristics and behaviors, genetic predispositions, and immunologic and other susceptibility-related factors that influence the likelihood or severity of disease. Host factors can be physiological, anatomical, genetic, behavioral, occupational, or constitutional. **Environmental factors** are external conditions other than the agent that contribute to the disease process. Environmental factors can be physical, biological, social, economic, or political in nature.

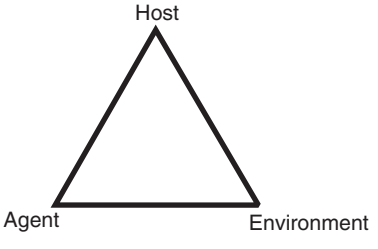


Figure 2.8 Agent, host, and environment.

Table 2.2 Types of disease-causing agents.

Biological	Chemical	Physical
Helminths (parasitic worms)	Nutritive (deficiencies and excesses)	Heat
Protozoan	Poisons	Light
Fungi	Drugs	Radiation
Bacteria	Allergens	Noise
Rickettsia		Vibration
Virus		Trauma
Prions		

The **sexual transmission of HIV** in a population can be viewed in terms of agent, host, and environmental determinants (Figure 2.9). Agent factors that influence HIV transmission include the prevalence of the agent in the environment and the phenotype of the agent. Examples of host factors include the coexistence of reproductive tract infections (especially genital ulcers), availability of antiretroviral therapies that decrease the HIV load in the population, prevalence of risky sexual behaviors, and use of condoms. Environmental factors include the rate of sexual partner exchange, presence of unregulated commercial sex facilities, presence of ''crack houses,'' sexual norms, and so on (Royce *et al.*, 1997).

Over time, an **epidemiologic homeostasis** may form as agent, host, and environmental factors reach equilibrium. When an element contributing to the epidemiologic

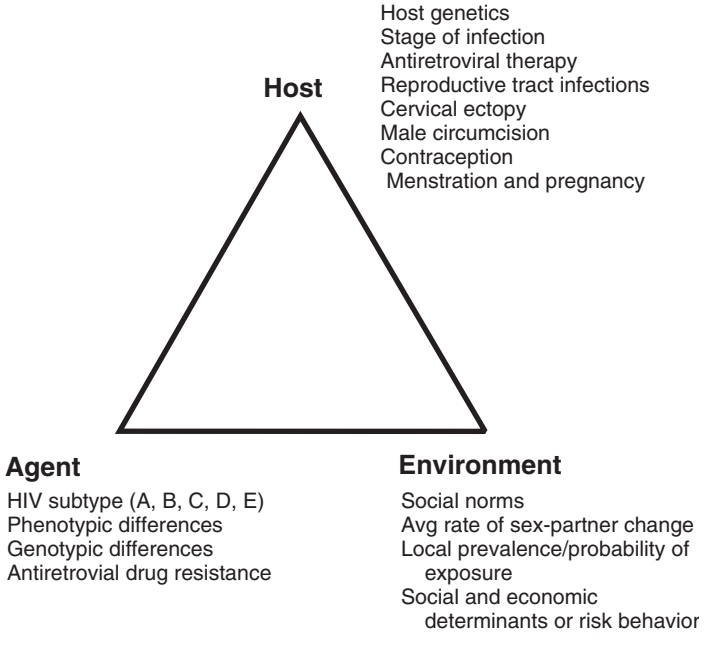


Figure 2.9 Agent, host, and environmental factors associated with the sexual transmission of HIV.

equilibrium is disturbed, the population may experience an increase or decrease in disease occurrence. For example, an epidemic may arise from any of the following:

- introduction of a new agent into the population
- increases in the ability of an agent to survive in the environment
- increases in an agent's ability to infect the host (**infectivity**)
- increases in the ability of the agent to cause disease once inside the host (**pathogenicity**)
- increases in the severity of the disease caused by the agent once it has established itself in the host (**virulence**)
- increases in the proportion of susceptibles in the population
- environmental changes that favor growth
- environmental changes that favor transmission of the agent
- environmental changes that compromise host resistance.

Causal forces can strengthen, weaken, or cancel-out each other, tipping the epidemiologic balance in favor of the host or in favor of the disease causing agent (Figure 2.10).

Homeostatic principles of agent, host, and environmental balance apply to infectious and noninfectious agents alike. As an example, consider the ecologic balance between agent, host, and environmental factors associated with **sulfur oxide air pollution and morbidity** (U.S. Department of Health, Education, and Welfare, 1967). In this example, high atmospheric levels of sulfur oxide pollution are traced to industrial pollution. Meteorologic conditions (e.g., climatic inversions) that favor retention of pollutants in the ecosphere have demonstrable effects on increasing morbidity and mortality, with the adverse effects of pollution concentrated in individuals with pre-existing cardiac and respiratory disease (Munn, 1970, p. 95). Thus, morbidity and mortality are linked to interdependencies between agent (e.g., sulfur dioxide pollution), host (compromised cardiopulmonary function), and environmental (meteorologic) conditions.

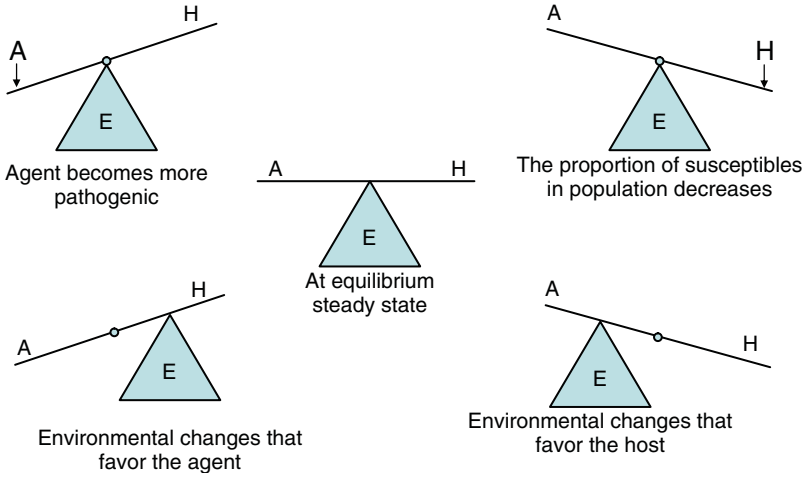


Figure 2.10 Agent, host, and environmental homeostasis and imbalance.

2.4 Causal inference

> The measures which are intended to prevent disease should be founded on a correct knowledge of its cause. For want of this knowledge, the efforts which have been made to oppose cholera have often had a contrary effect.
>
> *John Snow* (1855, p. 136)

In epidemiologic research we often observe a statistical association between an exposure and disease. The epidemiologist is aware that not all associations are causal and so must then address the issue of causality separately. The process by which we decide whether or not an association is causal is called **causal inference**. More formally, causal inference is the process of deriving cause-and-effect conclusions by reasoning from knowledge and factual evidence.

Before delving into the topic of how one decides whether an exposure is causal, it is helpful to acknowledge that there is no such thing as ultimate proof in empirical sciences (and epidemiology is indeed an empirical science). ''A statement in natural science can be made strong or even overwhelming . . . It is doubtful, however, if such proportions can ever be regarded as proved'' (Cornfield, 1954, p. 19). Thus, causal inferences in epidemiology require an enormous amount of patience and skill. Studies need to isolate various influences, and alternative explanations must be advanced and tested, bringing together various lines of evidence. No mechanical rules can be laid down. Delicate judgments are required. There is often ample opportunity for error and much room for legitimate disagreement. Although this is not often an easy process, it must be recognized that most of what we know about human health and disease comes from this type of process.

Types of decisions

The goal of causal inference is to create a framework for taking action in the face of varying degrees of uncertainty. In adopting a pragmatic framework, it is helpful to distinguish between **two different types of decisions**:

- decisions having to do with **scientific hypotheses**
- decisions **requiring prompt action**.

These two decision-making processes differ in several important respects. Inferences about scientific hypotheses are intentionally skeptical; alternative explanations and theories must be raised without restriction, even after reaching tentative conclusions. In contrast to the stringent level of skepticism required to address scientific hypotheses, public health, regulatory, and legal decisions cannot always afford the luxury of unrestrained scientific skepticism. A framework for making choices with the best evidence currently at hand is often required, for to decide not to make a decision may itself represent a costly choice.

Wynder (1994) notes that discoveries of many preventive measures pre-date discoveries of the specific mechanisms responsible for disease, often by many years (Table 2.3). In the same vein, there is evidence that the ''war on cancer'' initiated in the last quarter of the 20th century had been misdirected toward understanding carcinogenic mechanisms and discovering new treatments, when in fact applied preventive research might have met with better results (Bailar and Gornick, 1997).

Table 2.3 Discovery dates of measures to prevent selected diseases compared with the dates of identification of the causative agent.

Disease	Discoverer of preventive measure	Year preventive measure discovered	Year agent was discovered	Causative or (preventive) agent	Discoverer of agent
Scurvy	J. Lind	1753	1928	[Ascorbic acid]	A. Szent-Györgyi
Pellagra	G. Casal	1755	1924	[Niacin]	J. Goldberger
Scrotal cancer	P. Pott	1755	1933	Benzo[a]pyrene (chimney soot)	J.W. Cook
Smallpox	E. Jenner	1798	1958	*Orthopoxvirus*	F. Fenner
Puerperal fever	J. Semmelweis	1847	1879	*Streptococcus*	L. Pasteur
Cholera	J. Snow	1849	1893	*Vibrio cholerae*	R. Koch
Bladder cancer	L. Rehn	1895	1938	2-Napththylamine (aniline dye)	W.C. Harper
Yellow fever	W. Reed *et al*.	1901	1928	*Flavivirus*	A. Stokes
Oral cancer	A. Abbe	1915	1974	*N*′-nitrosonornicotine (chewing tobacco)	D. Hoffmann

Source: Wynder (1994, p. 548).

Thus, a utilitarian perspective provides for two complementary types of inference: those having to do with activities requiring immediate attention, and those having to do with scientific knowledge. Both processes must remain open to self-correction, although the former is more lenient in allowing for tentative conclusions based on incomplete understandings.

Report of the Advisory Committee to the U.S. Surgeon General, 1964

Important debates over how best to infer causality from epidemiologic data intensified in the years following World War II. Many of these debates centered around the role of cigarettes in the development of lung cancer. In 1964, the Surgeon General of the United States convened a panel of scientists to advise him on this issue. This panel wrote a landmark report that established standards to address the use of epidemiologic data (U.S. Department of Health, Education, and Welfare, 1964). Acceptance of these standards and constructs has provided a framework for epidemiologic debates ever since. Some of the key constructs established by this report are:

- When coupled with clinical, pathological, and experimental data, results from epidemiologic studies can provide the basis upon which judgments of causality may be made.
- In carrying out epidemiologic studies, many variables must be considered. In addition, the results of multiple investigations must be considered to determine first whether an association actually exists between the attribute or agent and disease.
- If it is shown that an association actually exists, then the question is asked: ‘‘Does the association have a causal significance?’’
- Statistical methods alone cannot establish proof of a causal relationship in an association. The causal significance of an association is a matter of judgment that goes beyond any statement of statistical probability.

- To judge causal significance, a number of criteria must be considered. No single one of these criteria is an all-sufficient basis for judgment. Criteria include:
  1 consistency of the association across studies
  2 strength of the association
  3 specificity of the association
  4 establishing a proper temporal relationship for the association
  5 coherence of the association.

At the time, these points helped to change the way in which the scientific community thought about epidemiologic data. Although the above causal criteria were not innovation of the committee, having been developed gradually over time, their value cannot be overlooked.

Within a year of the ''Surgeon General's Report,'' the British scientist A. Bradford Hill delivered his now classic paper called ''The environmental and disease: Association or causation?'' to the section of occupational medicine of the Royal Society of Medicine.

Hill's framework for causal inference

Bradford Hill's 1965 paper presented a framework for considering whether observed associations derived from epidemiologic may be considered to be causal. In this framework, Hill presents these nine elements:

1 strength
2 consistency
3 specificity
4 temporality
5 biological gradient
6 biological plausibility
7 coherence
8 experimentation
9 analogy.

Element 1 (strength) holds that strong associations provide firmer evidence of causality than do weak ones, and that the most direct measure of the strength of an association is found in the form of the ratio of two incidences (i.e., the relative risk).[b] According to this criterion, the larger the relative risk, the stronger the evidence for causality. The basis of this ''strength argument'' lies in the difficulty of explaining away a strong association as an artifact of an undiscovered extraneous factor causing a spurious association. In contrast, explaining away a small association with an unconsidered confounding factor is much more likely. Note, however, that the converse argument—that weak associations provide evidence that the association is noncausal—is not pertinent. ''We must not be too ready to dismiss a cause-and-effect hypothesis merely on the grounds that the observed association appears slight. There are many occasions in medicine when this is in truth so'' (Hill, 1965, p. 296).

Element 2 (consistency) suggests that it is important to demonstrate consistent associations across studies using diverse methods of study in different populations

[b] In the article ''On the origin of risk relativism'' (*Epidemiology*, 2010, **21**, 3–9), Poole discusses the historical context of using the relative risk as a measure of ''strength'' and questions the value of this practice.

under a variety of circumstances. The greater the number of consistent studies, the stronger the causal evidence. As an example, the data in Table 2.4 demonstrate highly consistent results between seven early cohort studies on smoking and lung cancer mortality. Note, however, that consistency alone is not sufficient to prove causation. Studies can be consistently wrong when they suffer from similar types of errors in design and interpretation.

Element 3 (specificity) holds that a causal factor that leads to a particular outcome provides stronger evidence than one that is connected to many. This criterion, however, should not be over-emphasized and is not always required. For example, smoking's propensity to contribute to many outcomes (cardiovascular disease, cancer, chronic lung disease, musculoskeletal disease, neurologic disease) cannot be used as an argument against its causal contribution to each, when in fact, there are specific biological mechanisms for each effect. Hill uses the example of milk as a non-specific cause of scarlet fever, diphtheria, tuberculosis, undulant fever, sore throat, dysentery, and typhoid fever, acting as a vehicle for each of the *Streptococcus* bacterial agents that cause each. Specificity of effect may therefore be interpreted in terms of the degree of association of the characteristic with the disease.

Element 4 (temporality) requires that exposure to the causal factor precede the onset of disease, and precede it by a reasonable amount of time. The importance of this criterion is self-evident. However, demonstration of accurate temporality is not always easy. The problem in sorting out the proper temporal sequence of events is especially troublesome when studying conditions with long latency and an insidious clinical onset. Consider the association between lead ingestion in children and impaired neuropsychological development. Lead encephalopathy is a clinical syndrome caused by ingestion of lead, with the greatest risk in young children exposed to decaying fragments of lead-based paint. Even though lead is a relatively common environmental contaminant capable of producing neurologic disease, it is not clear why some children develop lead encephalopathy while others do not. One theory suggests that children with behavioral problems and pica (a depraved or perverted appetite manifested by a hunger for substances not fit for consumption) are more likely to ingest lead. Pica is also associated with lower socioeconomic status and deficient care-giving, and this, too, can explain the association. Thus, the insidious onset of encephalitic symptoms and the complex interrelationships between environmental contamination with lead, pica, socioeconomic status, and behavioral disorders in children make it difficult to sort out the correct temporal relationships among these factors. Figure 2.11 (page 53) presents three different plausible temporal sequences that can explain the association between environmental lead and impaired neuropsychological development in children.

Element 5 (biological gradient) holds that an increase in the level, intensity, duration, or total exposure to an agent leads to progressive increases in risk. This is in keeping with the general toxicological principle of a quantal dose–response relationship, which states that the percentage of the population affected by a toxin increases as its dose is raised. In an epidemiologic dose–response relationship, the incidence of disease increases as the level of the risk factor is intensified. Examples of epidemiologic dose–response relationships are: the dose–response relationship between smoking and lung cancer (Figure 2.12, Table 2.4); the dose–response relation between serum cholesterol, systolic blood pressure, and coronary heart disease (Figure 2.13); and

Table 2.4 Cohort studies of smoking and lung cancer mortality.

Parameter	Doll and Hill (1956)	Hammond and Horn (1958)	Dorn (1958, 1959)	Dunn *et al*. (1960)	Dunn *et al*. (1964)	Best *et al*. (1961)	Hammond (1964)
Cohorts	British doctors	White men in nine states	US Veterans	California occupational groups	California American Legion members	Canadian pensioners and dependents	Men in 25 states
Number of people	34 000	188 000	248 000	67 000	60 000	78 000+	448 000
Age range	35–75+	50–59	30–75+	35–69	35–75+	35–75+	35–89
Months followed	120	44	78	About 48	About 24	72	About 22
Lung cancer deaths in study	129	448	535	139	98	221	414
Lung cancer deaths, nonsmokers	3	25	56	3	12	8	16
Cigs/day	Standardized mortality ratios (lung cancer)						
None	Referent	Referent	Referent	Referent	Referent	Referent	Referent
<10	4.4	5.8	5.2	(5)[a] 8.3	(≤20)[a] 4.2	8.4	—
10–20	10.8	7.3	9.4	(10)[a] 9.0	—	13.5	9.6[b]
21–39	—	15.9	18.1	(20)[a] 19.4	—	—	—
—	(21+)[a] 43.7	—	—	(30)[a] 25.1	(20+)[a] 7.4	15.1	—
40+	—	24.7	23.3	(40)[a] 28.7	—	—	—

[a]Indicates daily number of cigarettes smoked in which classes have been split or combined.
[b]Current smoker, amount not specified.
Source: U.S. Department of Health, Education, and Welfare (1964, pp. 83, 164).

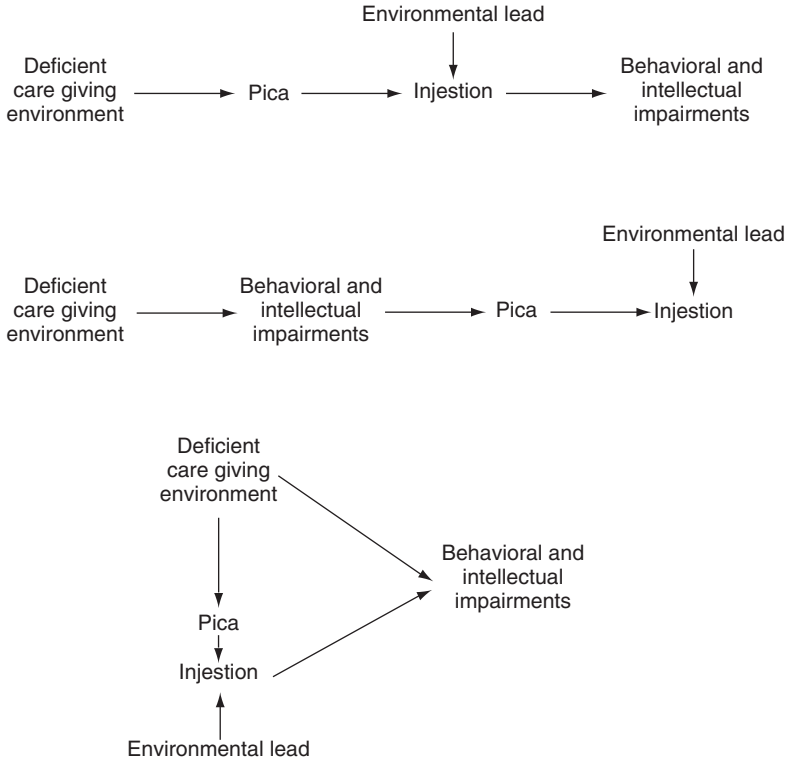


Figure 2.11 Three different plausible temporal sequences that can explain the association between lead encephalopathy and impaired psychological development in children.

the dose–response relation between oral contraceptive estrogen content and venous thromboembolic event risk (Figure 2.14).

Epidemiologic dose–response relationships come in different forms (e.g., linear, lognormal, ''U'' shaped, inverted ''U'' shaped), depending on the underlying pathophysiologic mechanism causing the elevations in risk. The type of dose–response relationship can have public health and regulatory implications. For instance, if there is a threshold response below which no further harm is done, further reduction in exposure is unwarranted; however, if risks are linearly related to cumulative dose throughout all potential levels of exposure, cumulative exposures must be minimized.

Element 6 (biological plausibility) holds that the association should be plausible with known biological facts. Statistical associations without sound biological reasoning may have little causal meaning. Consider the fact that most people die in bed. This undeniable statistical association is meaningless given common sense. This undeniable statistical association has little causal meaning given common sense and known biological fact. With this said, we must not be too ready to dismiss associations as noncausal simply because plausible explanations are unavailable. Biological plausibility is contingent on the current state of knowledge, and the current state of knowledge is constantly evolving.

Element 7 (coherence) holds that available evidence concerning the natural history, biology, and epidemiology of the disease should *stick together (cohere)* to

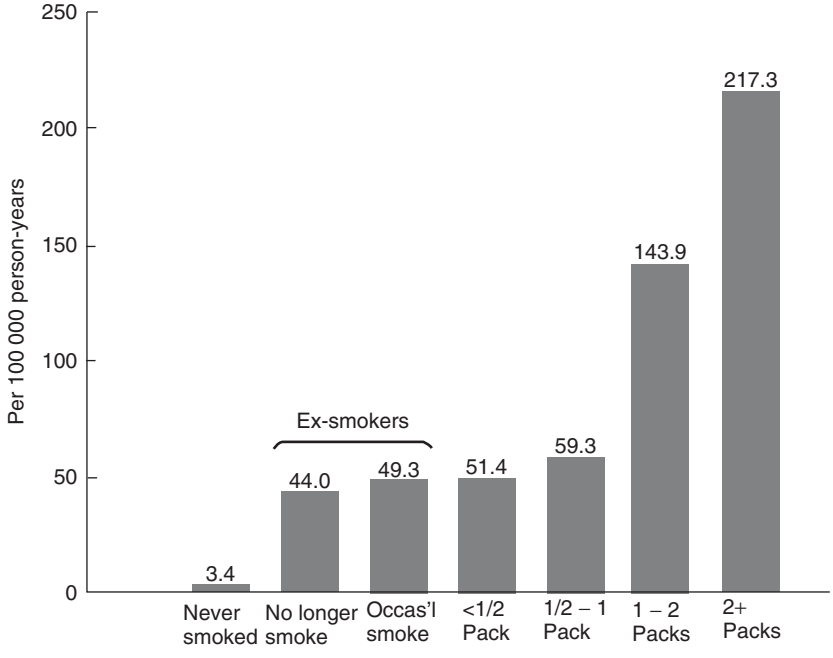


Figure 2.12 Age-adjusted death rates due to bronchogenic carcinoma exclusive of adenocarcinoma by current amount of cigarette smoking (Based on data in Hammond and Horn, 1958).

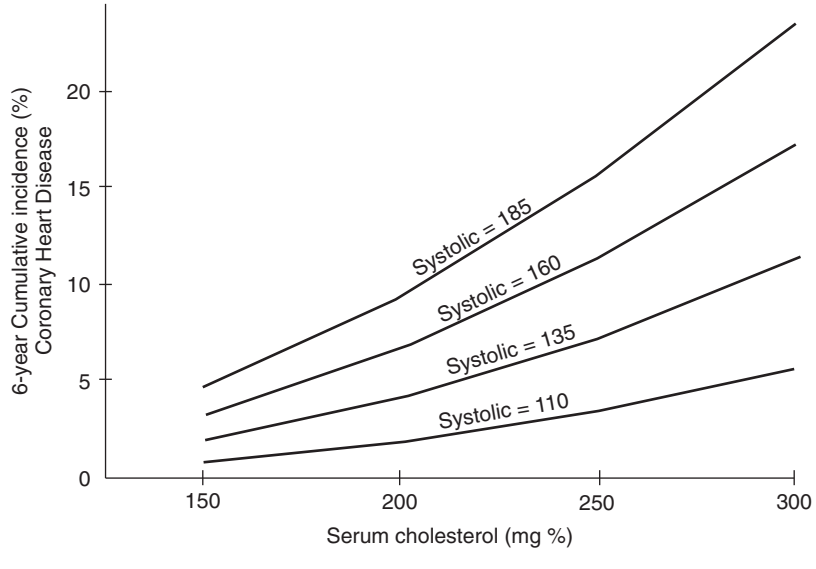


Figure 2.13 Six-year cumulative incidence of coronary heart disease according to serum cholesterol and systolic blood pressures, men 45–62 years old (Based on data in Kannel *et al.*, 1961).

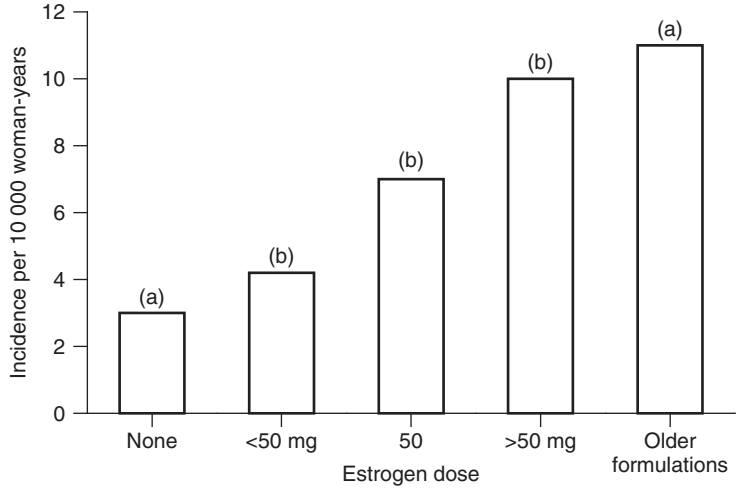


Figure 2.14 Oral contraceptive dose and rate of idiopathic deep venous thromboembolic disease (Based on data in (a) Stadel, 1981, p. 614; (b) Gerstman *et al*., 1991, p. 34).

form a unified whole. The proposed causal relation should neither conflict with nor contradict data derived from experimental, laboratory, and other epidemiologic sources. Consider, for example, that the rise of smoking in Western countries during the early and mid-20th century was accompanied by a corresponding increase in lung cancer mortality. This effect was more pronounced in men than in women, paralleling gender differences in the propensity to smoke. More recently, declines in lung cancer rates in men parallel recent declines in the prevalence of smoking in men but not in women (National Center for Health Statistics, 1995, p. 3). Experiments in animals support the presence of carcinogenic factors in cigarette smoke, and histopathologic evidence demonstrates the cytotoxic effect of smoking on the bronchial epithelium. These and other observations form a coherent whole in supporting the smoking and lung cancer causal hypothesis.

Element 8 (experimentation) addresses the need to support an observed association with experimental data from studies in human subjects and laboratory experiments. The strength of experimental trials in human subjects lies in their ability to randomize the factor being studied (see Chapter 6). Tests in the lab in the form of *in vitro* and *in vivo* experiments provide important support for causal mechanisms.

Element 9 (analogy) implies a similarity between things that are otherwise different. Although considered one of the weaker forms of inference, analogy can be useful in providing insights into the cause of a disease, especially during the early phases of an investigation. For example, an investigation to determine how the Lassa virus was spreading derived clues from the virus's morphological resemblance to the lymphocytic choriomeningitis virus and other arenaviruses that cause chronic infections in particular rodents. By analogy, the investigators were able to reason that some West African rodent may be susceptible to Lassa virus infection and may

infect humans through contaminated urine (Fraser, 1987). Thus, similar structures of otherwise dissimilar viruses led to clues about the source and transmission of the Lassa fever agent.

In considering Hill's inferential framework, it must be acknowledge that ***no*** *single element (except temporality) can be considered a necessary or indispensable condition in establishing causality*. Instead, causal inference relies on a compilation of judgments. Some uncertainty is inevitable. However, there is often a need to make casual judgments in the face of incomplete scientific knowledge. Hill reminds us of this responsibility with these parting words from his 1965 article (p.12):

> All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. This does not confer upon us a freedom to ignore the knowledge we already have, or to postpone action that it appears to demand at a given time . . . Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8.30 next day.

Philosophical considerations

Although a detailed discussion of the doctrines involved in scientific lines of inquiry is beyond the scope of this text, two key points must be emphasized. These are:

1. Scientists rely on the same method of reasoning common to all types of problem solving.
2. Induction and refutation have roles in epidemiologic practice.

1. Scientists rely on the same method of reasoning common to all types of problem solving. Although one may hear mention of ''the scientific method,'' it is not a method in the usual sense, since there are no orderly procedures and no rules of progression. Scientists rely on the same types of reasoning common to all types of problem solving. ''Scientific knowledge can only be an extension of common-sense knowledge'' (Popper, 1959, p. xxi); ''the scientific method, as far as it is a method, is nothing more than doing one's damnedest with one's mind, no holds barred'' (Bridgeman cited in Wallis and Roberts, 1962, p. 13). Astronomer Carl Sagan (1996) advises: ''We should not imagine that science is something erudite . . . The keypoint of science is criticism, debate, open inquiry, the willingness to systematize knowledge, to withhold belief until the evidence is compelling, and to listen seriously to criticism.'' Einstein said, ''If you want to know the essence of the scientific method, don't listen to what a scientist may tell you, watch what he does.'' In the end, ''science is only the Latin word meaning 'knowledge'.''

As a first-level introduction to problem solving, let's identify the following tools used during inquiry:

- **Observation**, in which the investigator observes what is happening, collects information, and studies facts relevant to the problem.
- **Hypothesis**, in which the investigator puts forth educated hunches or explanations for observed findings and facts.
- **Prediction**, in which anticipatory deductions based on hypotheses are put forward in testable ways.
- **Verification**, in which data are collected to test predictions.

Using these tools, inferences are established and tested as part of a continual and evolving process of learning. As conclusions unfold, they are reinforced through replication.

2. Induction and refutation have roles in epidemiology. Induction is the process of inferring a general law or principle from particular observations. **Refutation** is the process of putting forward and critically testing hypotheses through a process of falsification. Brief descriptions of induction and refutation follow.

Induction seeks to uncover the ''fabric of nature'' through observation of facts. The underlying assumption of induction is that phenomena which fall into regular patterns suggest more general statements about nature. The physicist Richard Feynman compared this process to watching a chess match without knowledge of the rules (Glashow, 1999):

> We can imagine that this complicated array of moving things that constitute the world is something like a great chess game played by the gods, and that we are observers of the game. We do not know what the rules of the game are; all we are allowed to do is to watch the playing. Of course, if we watch long enough, we may eventually catch onto a few of the rules. The rules of the game are what we mean by fundamental [laws of nature].

Induction is highly prone to error, however, because the sequences of past events are no guarantors of future occurrences. This is *The Problem of Induction*. The philosopher Bertrand Russell (1912/2008) explained the Problem of Induction in this memorable way:

> The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.

The formal argument against induction was made by David Hume when he wrote ''even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience'' (*A Treatise of Human Nature*, 1739–1740, Book I, Part III, Section XII). This is the logical fallacy of *post hoc ergo propter hoc* (Latin for ''after this therefore on account of this''). In a strictly logical sense, there is no reason to believe that what had been observed in the past will continue to occur in the future.

In recognizing the *Problem of Induction*, the influential 20th-century philosopher Karl Popper (1902–1994) placed central importance on the *Doctrine of Refutation*. In contrast to induction, refutation explicates the ''disproving'' of hypothetical statements as an essential component of scientific inference. Popper noted that statements about nature could not be proved in the affirmative but could be refuted through rigorous attempts to disprove falsifiable statements. By this method, failure to refute a hypothesis provides the best possible support of its verity. Because the absence of disproof is a demonstration of support for a hypothesis, the value of a given hypothesis depends on the degree to which it is ''disprovable.'' Thus, a theory is scientific to the extent that it is falsifiable. As the fictitious character Sherlock Holmes may have once remarked, ''when you have eliminated the impossible, whatever remains, however improbable, must be the truth.''

Scientific falsification has been explained in this way: Suppose two professors observe a flock of white swans around a campus pond. Being thoughtful academic types, they begin to wonder about the color of swans. The Inductionist professor induces that all swans are white based on all previous observations. In contrast, the Refutationist professor notes the observation and goes in search of non-white swans. If the Refutationist finds one non-white swan, the white swan hypothesis is revoked. Thus, there is a fundamental asymmetry of proof. No number of observations of white swans proves ''the white swan theory,'' whereas a strong refutation can disprove it.

While Popper's philosophy has had enormous practical benefits, it also has the potential to be misapplied. The absence of disproof is not proof, and like induction, falsification is limited by the senses. As might happen when a scientist returns with what is believed to be a non-white swan, he is often met with the response ''That's not a swan!'' Even the refutationist philosopher Popper recognized the limitations of this system of logic in the practice of science, admitting ''probability statements are . . . in some sense verifiable . . . '' (Susser, 1988, p. 193).

Thus,

> The true spirit of science is positive. The building of theory is art; it depends on imaginative synthesis, most often by inductive sifting, sometimes by a leap of the mind. The execution of tests (either falsification or verification) is craft; it depends on ingenuity and technique. The refutation of the theory of spontaneous generation was sealed by Louis Pasteur's verification of the positive role bacteria in fermentation (1862). Much earlier Spallanzini, and the Schulze, Schwann and others, had refuted the theory when they showed that under controlled conditions fermentation did not occur. Falsification was less successful here than verification because supporters of the theory could advance an endless series of alternative explanations. It is Pasteur's work that is remembered . . .
>
> *Susser*, 1988, pp. 195–196

Exercises

2.1 Select a specific disease or type of injury that interests you or with which you have some experience or knowledge. For example, if you are a pediatrician, you might select otitis media or some other common childhood disease; if you are a respiratory therapist, you might select asthma or COPD; if a relative of yours has diabetes, you may select this ailment (and so on). Then, after studying about the disease from a reliable medical source (e.g., *www.merckmanuals.com/home/* or *http://www.merck.com/pubs/mmanual/*), address each of the following questions:

- **(A)** Describe the spectrum of this disease and its range of clinical manifestations.
- **(B)** Identify host, agent, and environmental causal factors for this disease. Address both direct and indirect causes. Form these factors into a causal web.
- **(C)** List primary methods of prevention for this disease. Why, specifically, do you believe that these are primary methods of prevention, and not, say, secondary methods of prevention?
- **(D)** List secondary methods of prevention. Justify why you believe these are secondary and not primary or tertiary forms of prevention.

(E) List tertiary forms of prevention. Justify why you believe these are tertiary and not primary or secondary forms of prevention.

(F) Which of the above forms of prevention listed in parts (D) and (E) do you believe to be most effective. Justify your response.

2.2 *Multiple choice*. *Select the single best response*. This model of disease causation suggests that direct and indirect causes act together to form complex interrelations in a hierarchal manner.

(a) epidemiologic homeostasis
(b) component cause/causal pies
(c) ''agent, host, and environment''
(d) causal web.

2.3 *Multiple choice*. Smoking increases the risk of lung cancer. However, not all smokers develop lung cancer and some nonsmokers develop lung cancer. Therefore, smoking is a _____ causal factor for lung cancer.

(a) necessary
(b) sufficient
(c) contributing
(d) infectious.

2.4 *Multiple choice*. Mammographic screening is intended to detect breast cancer in its early stages. Therefore, mammographic screening is a form of _____ prevention.

(a) primary
(b) secondary
(c) tertiary
(d) mammographic screening is not a form of prevention.

2.5 *Multiple choice*. The incidence of the parasitic disease schistosomiasis can be reduced by ridding the environment of the species of snail that is an intermediate host in the life cycle of the *Schistosoma* species that causes the disease. Therefore, snail control is a form of _____ prevention.

(a) primary
(b) secondary
(c) tertiary
(d) snail control is not a form of prevention.

2.6 *Multiple choice*. A patient has chest pain and shortness of breath because of coronary heart disease and is treated for their accompanying hypertension. Is this a form of primary, secondary, or tertiary prevention?

(a) primary
(b) secondary
(c) tertiary
(d) treatment of hypertension is not a form of prevention.

2.7 *Multiple choice*. This term means the ability of an agent to cause disease, irrespective of the severity of the outcome.

(a) infectivity
(b) pathogenicity

(c) virulence
(d) effectiveness.

2.8 *Multiple choice*. This term specifically relates to the capacity of an agent to cause severe disease.
(a) infectivity
(b) pathogenicity
(c) virulence
(d) effectiveness.

2.9 *Multiple choice*. This refers to circumstances whereby only a small percentage of a problem is readily apparent.
(a) epidemiologic homeostasis
(b) the epidemiologic iceberg
(c) the causal web
(d) causal pies.

2.10 A causal interaction occurs when:
(a) two or more factors acting together bring about an effect
(b) the period between exposure to the agent and first symptoms of disease is brief
(c) there is a balancing of agent, host, and environment maintaining a constant rate of disease in the population
(d) the outcome becomes inevitable.

2.11 *Matching*. Match each of these terms with the definitions provided below: Causal-web; Component Cause; Infectivity; Necessary Causal Factor; Pathogenicity; Sufficient Causal Constellation; Virulence.
Definitions:
(a) Related to the *severity* of disease.
(b) A set of factors that makes disease inevitable in an individual.
(c) An antecedent factor that contributes to the disease process in some individuals.
(d) The ability of a communicable agent to enter and establish itself in a host.
(e) The causal model that links direct and indirect causes of disease through an intertwined hierarchy.
(f) The ability of an agent to cause disease.
(g) A factor that is always present for a given disease; the disease cannot occur without it.

2.12 *Matching*. Match the descriptions of each of Hill's considerations with one of these brief descriptive labels: Strength; Consistency; Specificity; Temporality; Biological gradient; Plausibility; Coherence; Experimentation; Analogy.
Descriptions of causal considerations:
(a) This criterion holds that all available clinical, experimental, and observational evidence should ''stick together'' in the argument for causation.
(b) This criterion holds that an increase in the level, intensity, duration, or total level of exposure leads to progressive increases in the magnitude of risk.
(c) This criterion holds that an association is explainable in terms of known biological fact.

(d) This criterion requires that exposure to the causal factor precedes the onset of disease by a reasonable amount of time.

(e) This criterion requires supporting evidence from community and clinical trials, *in vitro* laboratory experiments, and animal models.

(f) This criterion is based on similarities from otherwise dissimilar sources.

(g) This criterion holds that the cause should lead to only one disease and that the disease should result from this single cause only.

(h) This criterion holds that diverse methods of study carried out in different populations under a variety of circumstances by different investigators provide similar results.

(i) This criterion holds that strong associations provide firmer evidence of causality than do weak ones.

2.13 *Matching*. The association between oral contraceptives and cardiovascular disease has been the subject of considerable debate. Indicate which of Hill's causal considerations are addressed by each of the statements below. Use these labels in tagging the appropriate consideration: Strength; Consistency; Specificity; Temporality; Biological gradient; Plausibility; Coherence; Experimentation; Analogy.

(a) The risk of cardiovascular disease increases with increasing the estrogen dose of the oral contraceptive formulation.

(b) Studies have shown that oral contraceptives cause endothelial proliferation, decrease the rate of venous blood flow, and increase the coagulability of blood by altering platelet function, coagulation factors, and fibrinolytic activity.

(c) The relative risk of oral contraceptive use and mortality from all circulatory disease in the 1970s was approximately 4. (A relative risk of 4 means that the women taking oral contraceptives have a rate of cardiovascular disease that is 4 times that of women who do not, all other things being equal.)

(d) Nearly all studies completed to date have demonstrated a positive association between oral contraceptive use and cardiovascular disease risk.

(e) Other steroidal sex hormones, such as testosterone, have known effects on cardiovascular disease risk.

(f) Altered parameters of hemostasis are measurable soon after oral contraceptives are begun. These alterations return to baseline within a month of discontinuing oral contraceptives.

2.14 *True or false?* Hill's criterion for ''consistency'' holds that the exposure will always lead to the disease. If the statement is false, state why it is false and supply the information that would make the statement true. *If false, modify the statement to make it true*.

Review questions

Sections 2.1–2.3

R.2.1 List the four stages in the natural history of disease.

R.2.2 What events mark the beginning of each stage of disease?

R.2.3 What is the goal of primary prevention? . . . of secondary prevention? . . . of tertiary prevention?

R.2.4 What occurs during the incubation period of an infectious disease?

R.2.5 Provide a synonym for ''incubation period.''

R.2.6 Provide an example of an infectious disease with a long incubation period.

R.2.7 Is mammography a form of primary, secondary, or tertiary prevention? Explain your response.

R.2.8 Discuss the natural history of HIV/AIDs and its relevance to AIDs prevention.

R.2.9 This term is used to describe the broad range of clinical manifestations for an ailment.

R.2.10 This metaphor refers to the circumstance in which only a small percentage of a problem is readily apparent.

R.2.11 Define ''cause.''

R.2.12 What is causal interaction?

R.2.13 Is the measles virus a necessary cause of measles? Is it a sufficient cause? Explain your responses.

R.2.14 *When* is a causal mechanism complete?

R.2.15 Suppose a disease mechanism has three causal components: D, E, and F. What is the causal complement to D?

R.2.16 Refer to the prior question. What is the causal complement to (D+F)?

R.2.17 Phenylketonuria is a rare disorder in which a person is born without the ability to properly break down an amino acid called phenylalanine due to an enzyme deficiency. Symptoms occur only when a susceptible individual is exposed to phenylalanine in the diet. Is phenylketonuria a genetic disease or an environmental disease? Explain your response.

R.2.18 List at least four component causes of hip fractures in the elderly.

R.2.19 How does a direct cause of disease differ from an indirect cause?

R.2.20 List the three general categories of pathogenic agents.

R.2.21 List four types of chemical agents of disease.

R.2.22 List five types of physical agents of disease.

R.2.23 Differentiate between infectivity, pathogenicity, and virulence.

R.2.24 Describe the concept of *epidemiologic homeostasis*.

Section 2.4 *Causal inference*

R.2.25 What is causal inference?

R.2.26 Why do we base preventive measures on knowledge of causal mechanisms?

R.2.27 When is it necessary to go forward with an intervention before full knowledge of a causal mechanism is fully complete?

R.2.28 Provide an example in which the discovery of an effective measure to prevent disease pre-dated the discovery of its causal mechanism.

R.2.29 In what year was the initial Surgeon General's Report on Smoking and Health published?

R.2.30 Why does association not necessarily equate with causation?

R.2.31 True of false? Statistical methods can establish proof of causation.

R.2.32 By convention, what is the most direct measure of the strength of an association?

R.2.33 Why do strong associations provide stronger evidence of cause than weak associations?

R.2.34 Are weak associations indicative of a noncausal relationship?

R.2.35 If multiple studies consistently show the same results, is this proof of causation? Explain your response.

R.2.36 True of false? Establishing the proper temporal relation between the cause and its effect is mandatory when establishing cause. Explain your response.

R.2.37 How does the element of coherence differ from that of plausibility? (Tough one.)

R.2.38 Early in the AIDS epidemic, before HIV was discovered (circa 1983), epidemiologists realized groups at high risk of HIV groups shared characteristics with groups at high risk of Hepatitis B. This suggested the diseases were spread by similar mechanisms. Which of Hill's casual elements is being addressed by this argument.

R.2.39 A study on micro-nutrient and Alzheimer's disease found progressively lower risks of brain atrophy with increasing levels of folic acid consumption (Snowdon *et al.*, 2000). Which of Hill's causal criteria is addressed by this statement?

R.2.40 List short titles for each of the elements in Hill's causal framework.

Optional section on Philosophical considerations

R.2.41 Is there such thing as ultimate proof in empirical sciences? Explain your response.

R.2.42 Two types of epidemiologic decisions are (1) those having to do with scientific hypotheses and (2) those having to do with public health interventions. To what extents do each of these types of decisions require skepticism.

R.2.43 Describe the philosophical Problem of Induction.

R.2.44 True or false? According to refutationist theory, the value of a scientific hypothesis depends on the degree to which it can be disproved. Discuss your response.

R.2.45 Can we ever prove that all swans are white? Explain your response.

R.2.46 Describe the asymmetry of positive and negative proof.

References

Bailar III,, J.C., and Gornick, H.L. (1997) Cancer undefeated. *New England Journal of Medicine*, **336**, 1569–1574.

Benenson, A.S. (ed.) (1990) *Control of Communicable Diseases in Man*, 15th edn, American Public Health Association, Washington D.C.

Best, E.W.R., Josie, G.H., and Walker, C.B.A. (1961) Canadian study of mortality in relation to smoking habits, a preliminary report. *Canadian Journal of Public Health*, **52**, 99–106.

Case, R.A.M., Hosker, M.E., McDonald, D.B., and Pearson, J.T. (1954) Tumors of the urinary bladder in workman engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. *British Journal of Industrial Medicine*, **11**, 75–104.

Cobb, S., Miller, M., and Wald, N. (1959) On the estimation of the incubation period in malignant disease. *Journal of Chronic Diseases*, **9**, 385–393.

Cornfield, J. (1954) Statistical relationships and proof in medicine. *American Statistician*, **8**, 11–13.

Cotton, D.J. (1995) The natural history of HIV infection: implications for clinical trials, in *AIDS Clinical Trials*, (eds D.M. Finkelstein and D.A. Schoenfeld), Wiley-Liss.

Doll, R., and Hill, A.B. (1956) Lung cancer and other causes of death in relation to smoking. *British Medical Journal*, **2**, 1071–1081.

Dorn, H.F. (1958) The mortality of smokers and non-smokers. *Proceedings of the Social Statistics Section. American Statistical Association. Social Statistics Section*, 34–71. (As cited in U.S. Department of Health, Education, and Welfare, 1964.)

Dorn, H. F. (1959) Tobacco consumption and mortality from cancer and other diseases. *Public Health Reports*, **74**, 581–593.

Dunn, J.E., Jr., Linden, G., and Breslow, L. (1960) Lung cancer mortality experience of men in certain occupations in California. *American Journal of Public Health*, **50**, 1475–1487.

Dunn, J.E., Jr., Linden, G., and Breslow, L. (1964) California State Department of Public Health. *Special Report to the Surgeon General's Advisory Committee on Smoking and Health*. (Data published in U.S. Department of Health, Education and Welfare, 1964.)

Fraser, D.W. (1987) Epidemiology as a liberal art. *New England Journal of Medicine*, **316**, 309–314.

Gerstman, B.B., Piper, J.M., Tomita, D.K., Ferguson, W.J., Stadel, B.V., and Lundin, F.E. (1991) Oral contraceptive estrogen dose and the risk of deep venous thromboembolic disease. *American Journal of Epidemiology*, **133**, 32–37.

Glashow, S.L. (1999) Can science save the world? Centro Congressi Cariplo. Available: *http://www.hypothesis.it/nobel/nobel99/eng/pro/pro_2_1_1.htm* (accessed 8 April 2002).

Hammond, E.C. (1964) Prospective study of 1,085,000 men and women in 25 of the United States aged 35–89. *Special report to the Surgeon General's Advisory Committee on Smoking and Health*. (Data published in U.S. Department of Health, Education and Welfare, 1964.)

Hammond, E.C., and Horn, D. (1958) Smoking and death rates—report of forty-four months of follow-up of 187,783 men. Part I. Total mortality. Part II. Death rates by cause. *JAMA*, **166**, 1159–1172, 1294–1308.

Hill, A.B. (1965) The environmental and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, **58**, 295–300.

Hogben, L. (1933) *Nature and Nurture*, Williams and Norgate, London.

Hume, D. (1739–40) *A Treatise of Human Nature*, McMaster University Archive. Available: *http://www.ets.uidaho.edu/mickelsen/ToC/hume%20treatise%20ToC.htm* (accessed 2 June 2002).

Hume, D. (1772) *An Enquiry Concerning Human Understanding* (1993 web reprint of the Harvard Classic published in 1910). Available: *http://eserver.org/18th/hume-enquiry.html*.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., and Stokes, J. (1961) Factors of risk in the development of coronary heart disease—six year follow-up experience. *Annals of Internal Medicine*, **55**, 33–50.

Last, J.M. (1963) The iceberg: 'completing the clinical picture' in general practice. *The Lancet*, 6 July, 28–31.

Millard, A.V. (1994) A causal model of high rates of child mortality. *Social Science and Medicine*, **38**, 253–268.

Munn, R.E. (1970) *Air Quality for Sulfur Oxides*, Academic Press, New York.

National Center for Health Statistics (1995) *Health, United States, 1994*, Public Health Service, Hyattsville, MD.

Poole, C. (2010) On the origin of risk relativism. *Epidemiology*, **21** (1), 3–9.

Popper, K. (1959) *The Logic of Scientific Discovery*, Routledge Classics, London.

Rothman, K.J. (1976) Causes. *American Journal of Epidemiology*, **104**, 587–592.

Rothman, K.J. (1981) Induction and latent periods. *American Journal of Epidemiology*, **114** (2), 253–259.

Rothman, K.J., and Greenland, S. (1998) *Modern Epidemiology*, 2nd edn, Lippincott-Raven, Philadelphia.

Royce, R.A., Sena, A., Cates, W., Jr., and Cohen, M.S. (1997) Sexual transmission of HIV. *New England Journal of Medicine*, **336**, 1072–1078.

Russell, B. (2008) *The Problems of Philosophy*. Arc Manor, Rockville, MD (original work published 1912).

Sagan, C. (1996, December 20) Carl Sagan Obituary. Bob Evans (Host) *Morning Edition*, National Public Radio, Washington D.C.

Savitz, D.A. (1997) The alternative to epidemiologic theory: Whatever works. *Epidemiology*, **8**, 210–212.

Snow, J. (1855) *On the Mode of Communication of Cholera*, John Churchill, London. Reprinted (1936) in *Snow on Cholera*, The Commonwealth Fund, New York.

Snowdon, D.A., Tully, C.L., Smith, C.D., Riley, K.P., and Markesbery, W.R. (2000) Serum folate and the severity of atrophy of the neocortex in Alzheimer disease: findings from the Nun study. *American Journal of Clinical Nutrition*, **71** (4), 993–998.

Stadel, B.V (1981) Oral contraceptives and cardiovascular disease. *New England Journal of Medicine*, **305**, 612–618, 672–677.

Susser, M. (1988) Falsification, verification and causal inference in epidemiology: Reconsideration in the light of Sir Karl Popper's philosophy, in *Causal Inference* (ed. K.J. Rothman), Epidemiology Resources, Inc., Chestnut Hill, MA, pp. 33–57 [Rejoinder: Rational science versus a system of logic, in *Causal Inference* (ed. K.J. Rothman) (pp. 189–199)].

U.S. Department of Health, Education, and Welfare (1964) Smoking and health. Report of the Advisory Committee to the Surgeon General of the Public Health Service. USPHS Publication No. 1103, U.S. Government Printing Office, Washington D.C. (Also published by D. Van Nostrand Company, Princeton, New Jersey.)

U.S. Department of Health, Education, and Welfare, National Center for Air Pollution Control. (1967) Air quality criteria for sulfur oxides. Public Health Service, Washington, D.C. (for sale by the Supt. of Docs., U.S. Government Printing Office).

Wallis, W.A., and Roberts, H.V (1962) *The Nature of Statistics*, Free Press, New York.

Weiss, H.B., Friedman, D.I., and Coben, J.H. (1998) Incidence of dog bite injuries treated in emergency departments. *JAMA*, **279**, 51–53.

Wynder, E.L. (1994) Invited commentary: Studies in mechanism and prevention. Striking a proper balance. *American Journal of Epidemiology*, **139**, 547–549.

CHAPTER 3
Epidemiologic Measures

3.1 Measures of disease frequency
- Background
- Incidence proportion (risk)
- Incidence rate (incidence density)
- Prevalence

3.2 Measures of association
- Background
- Absolute versus relative comparisons
- Absolute measures of effect
- Relative measures of effect
- Odds ratios
- Relation between the *RR* and *RD*

3.3 Measures of potential impact
- Attributable fraction in the population
- Attributable fraction in exposed cases
- Preventable fraction

3.4 Rate adjustment
- Background
- Direct adjustment
- Indirect adjustment
- Adjustment for multiple factors
- Section summary
- Notation used in Section 3.4

Exercises

Review questions

References

Addendum: additional mathematical details

Epidemiologic measures are used to quantify (a) the frequency or events and conditions in populations, (b) the effects of an exposure, and (c) the potential impact of an intervention. We start by considering measures of disease frequency.

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

3.1 Measures of disease frequency

Background

Measures of disease frequency quantify how often a disease or condition occurs within a given population. Thus, measures of disease frequency are also called **measures of occurrence**.

The three main measures of disease frequency are:

- incidence proportion (risk)
- incidence rate (incidence density)
- prevalence.

All three of these measures of disease frequency are types of **ratios** consisting of a numerator and denominator. The **numerator** of each measure of disease frequency is some type of count of cases. The **denominator** is a measure of population size or ''person-time.'' As you learn each of these measures of disease frequency, pay careful attention to the similarities and differences in their numerators and denominators.

An important consideration when studying disease frequency is whether the population being studied is closed or opened.

1 **Closed populations** are also called **cohorts**. They gain no new members after they are established and lose members only when members die or are no longer at risk of becoming a case for whatever reason. Thus, they begin with a certain number of individuals and shrink over time as mortality takes its toll and individuals are removed from risk. Figure 3.1 illustrates the survival experience of a birth cohort over time. Note how the cohort shrinks and ages over time.

2 In contrast to closed populations, **open populations** add new members through birth and immigration, and lose members through emigration and death. Over time, open populations may grow, remain the same size, or shrink, depending on their rate of inflow or outflow (Figures 3.2 and 3.3). An open population that is in a steady state—so, for example, as one person dies a new individual is born into the population and another ages up to replace the death—is said to be **stationary**.

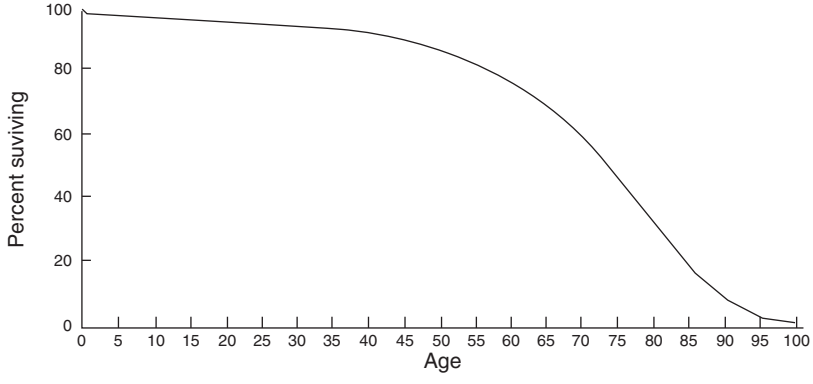


Figure 3.1 Percent surviving in a birth cohort (closed population) over time.

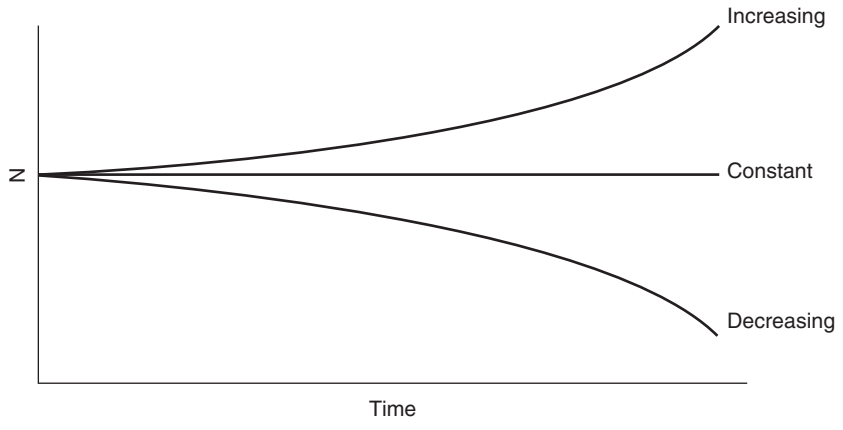


Figure 3.2 An open population may grow, stay constant in size, or decrease is size over time.

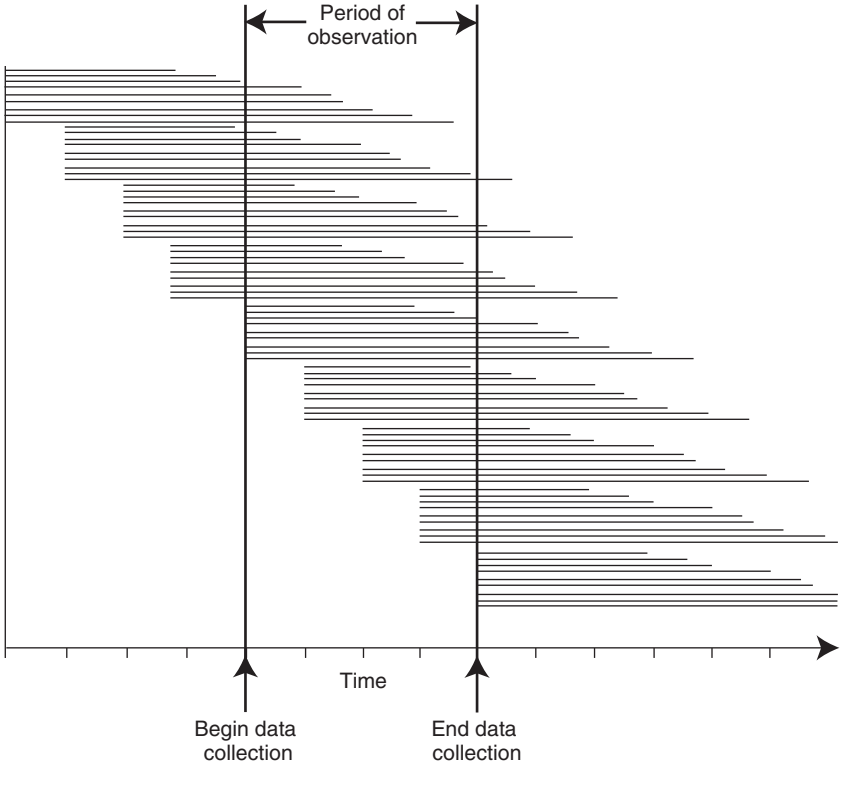


Figure 3.3 An open population in a steady-state is said to be stationary.

Incidence proportion (risk)

The **incidence proportion** (**cumulative incidence, average risk**) of a disease is a longitudinal measure of disease occurrence in which the numerator consists of the number of disease onsets that occurred during the period of observation and the denominator consists of the number of individuals at risk in the closed population as of the beginning of follow-up:

$$\text{Incidence proportion} = \frac{\text{no. of disease onsets}}{\text{no. of individuals at risk at beginning of follow-up}} \quad (3.1)$$

Note that incidence proportions can only be measured in cohorts, and cannot be calculated in open populations. Also note that the denominator includes only individuals at risk of developing the condition being studied and therefore excludes individuals who are not capable of developing the condition under consideration. For example, in studying uterine cancer, the denominator excludes women who had already experienced uterine cancer, women with prior hysterectomies, and (of course) men, because these individuals are not capable of developing the condition being studied.

Interpretation: An incidence proportion is the **average risk** of developing the condition under consideration for the period of observation. Therefore, the terms *incidence proportion* and *risk* are used interchangeably in epidemiology. In addition, since the incidence proportion represents an accumulation of new cases over time, it is also referred to as *cumulative incidence*.

To interpret an incidence proportion properly, the length of the time at risk must be specified. In addition, characteristics of the population should be made clear. Consider, for example, the incidence proportion (risk) of breast cancer in American women. The lifetime risk for this outcome is 12% (1 in 8). In contrast, the risk in women between the ages of 60 and 69 is 3.5% (1 in 29). Finally, the risk between ages 50 and 59 is 2.4% (1 in 42). Our understanding of population characteristics and the length of follow-up should temper our interpretation of an incidence proportion.

Illustrative Example 3.1 Cohort $n = 5$ (incidence proportion)

Figure 3.4 represents the experience of a cohort consisting of five people followed for up to ten years. Each horizontal line in the schematic represents the experience of an individual. Two disease onsets occurred during the ten years of follow-up. Therefore,

the ten-year incidence proportion = $\frac{2 \cancel{\text{ people}}}{5 \cancel{\text{ people}}} = 0.40$ or 40%.[a]

Illustrative Example 3.2 Uterine cancer (incidence proportion)

Let us consider a study that recruits 1000 women for a study of uterine cancer. Upon initial examination, the investigators discover that 100 of the potential study subjects had either already experienced uterine cancer or had a prior hysterectomy. This leaves 900 individuals at risk for uterine cancer. This cohort is followed for five years during which time 45 study subjects develop the disease. Thus, the five-year incidence proportion (risk) of uterine cancer = $\frac{45 \cancel{\text{ persons}}}{900 \cancel{\text{ persons}}} = 0.05$ or 5%.

[a] Notice that the ''people units'' in the numerator and denominator of the incidence proportion calculation cancel-out upon division leaving a unit-free number.

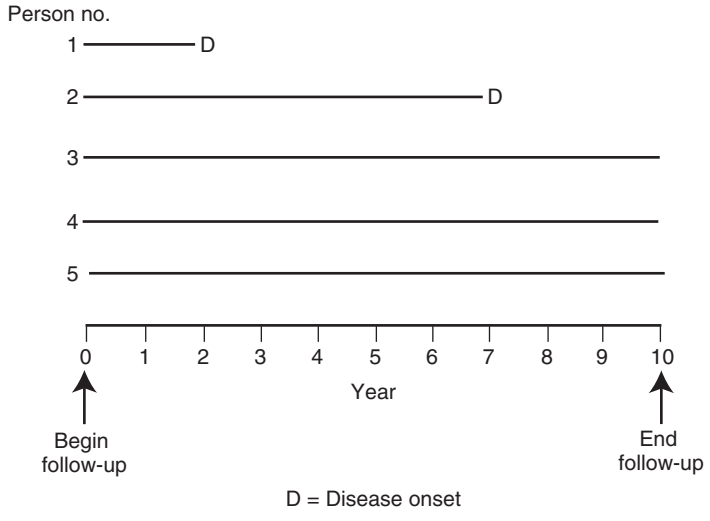


Figure 3.4 Schematic for Illustrative Example 3.1.

Incidence rate (incidence density)

The **incidence rate** (**incidence density**) of a disease is the number of disease onsets divided by the sum of person-time in the population:

$$\text{Incidence rate} = \frac{\text{no. of disease onset}}{\text{sum of person-time at risk}} \tag{3.2}$$

A **person-time unit** is the amount of time a person is observed during the study. One person observed for one year contributes one person-year to the denominator. One person observed for two years accounts for two person-years. Two people observed for one year each also accounts for two person-years (and so on). Note that person-time is counted only when a person is at risk of being detected as a case. Person-time is no longer counted after: (a) the person develops the disease under investigation, (b) the person withdraws from the study, or (c) the study ends.

Interpretation: We may interpret incidence rates in several compatible ways. Firstly, incidence rates represent the *speed*, *rapidity*, *density* or *intensity* at which populations are expected to *generate* cases. For example, a rate of 5 per 100 person-years is expected, on the average, to generate 5 cases in 100 people followed for one year.

Secondly, incidence rates reflect the incidence proportion (risk) of the disease when the disease is ''rare''[b] according to the formula: Risk $\approx$ Rate $\times$ Time. For example, a rate of 1 per 100 person-years over a one-year period corresponds to a one-year risk $\approx$ Rate $\times$ Time = (0.01 per person $-$ year) $\times$ (1 year) = 0.01 or 1%

Thirdly, the incidence rate in a population is related to its survival experience. Figure 3.5, which is an adaptation of Figure 3.4, is intended to give the reader insight into this relationship. Note that the *area under the curve* in this diagram is equivalent to the person-time in the cohort. (See Chapter 17 for additional information about the relationship between the probability of survival and rates of occurrence.)

[b] For the current purpose ''rare'' is defined as having an incidence proportion of 5% or less.

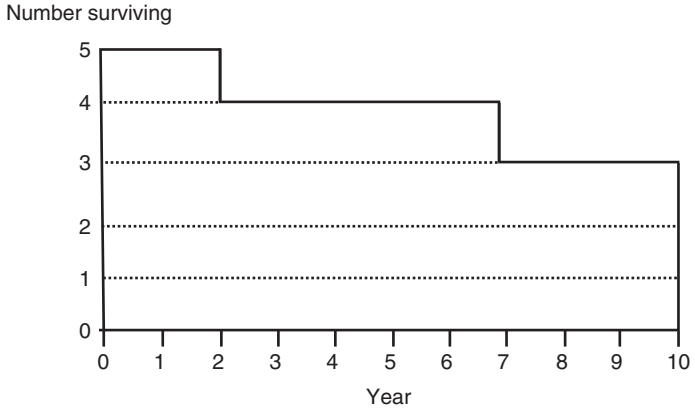


Figure 3.5 Survival experience of the cohort represented in Figure 3.4.

Illustrative Example 3.3 Cohort *n* = 5 (incidence rate, counting person-time)

Let us reconsider the data in Figure 3.4. In this schematic, person 1 contributes 2 person-years at risk, person 2 contributes 7 person-years at risk, and persons 3, 4, and 5 contribute 10 person-years at risk each. Thus, the sum of person-time in the cohort = 2 + 7 + 10 + 10 + 10 = 39 person-years. Two incidents of disease occur during the period of observation. Therefore, the incidence rate $= \frac{2 \text{ persons}}{39 \text{ person-years}} = 0.0513$ per year or, equivalently, 5.13 per 100 person-years.[c]

When information is not available on individual follow-up time in a cohort, we can estimate the person-time in the cohort with this equation:

$$\Sigma \text{ ''Person-time''} \approx \text{number of individuals at risk} \times \text{duration of ''follow-up''} \quad (3.3)$$

Illustrative Example 3.4 Incidence rate, approximating person-time

In Illustrative Example 3.2 we considered a cohort of 900 individuals at risk followed for 5 years each, during which time 45 incident cases emerged. Using Equation (3.3), Σ Person-time $\approx$ (no. of individuals at risk) $\times$ (duration of follow-up) = 900 persons $\times$ 5 years = 4500 person-years. Thus, the incidence rate $\approx \frac{45}{4500 \text{ person-years}} = 0.010\,0$ per person-year or 1.00 per 100 person-years.

Actuarial adjustment: Note that the 45 individuals in this example that developed disease did so at some time *during* the 5 years of follow-up. Therefore, most did not contribute the full 5 person-years at risk. (After developing the condition they are no longer at risk.) We can adjust for this phenomenon by assuming that the average time of onset of disease was half-way through the follow-up period. Thus, we assume each case contributed half of the 5 years at risk, or 2.5 years each. This method is called the *actuarial adjustment*. Therefore, a slightly more accurate estimate for the incidence rate in this cohort is $\frac{45}{(855 \text{ non-cases} \times 5 \text{ years}) + (45 \text{ cases} \times 2.5 \text{ years})} = 0.0103$ per person-year or 1.03 per 100 person-years.[d] Notice that this adjustment did not have a large effect because the incidence of the outcome is relatively modest.

[c] It is common to use a population multiplier when reporting rates. To report a rate ''per *m* person-years'' simply multiple that rate by *m*.
[d] The actuarial adjustment makes little difference when the disease is rare.

Up until this point we have considered rates only in closed populations (cohorts). Rates can also be estimated in **open populations** with this formula:

$$\text{Incidence rate} = \frac{\text{no. of disease onsets}}{(\text{average population size}) \times (\text{duration of observation})} \quad (3.4)$$

In open populations that are rapidly increasing or decreasing in size, it is common to use the population size mid-way through the period of observation as an estimate of the average population size.

Illustrative Example 3.5 Mortality (incidence rate in an open population)

In 2006, the United States had a mid-year population size of 299 398 000 residents. There were 2 426 264 deaths in that year. Therefore, the mortality rate $= \frac{2\,426\,264}{299\,398\,000 \text{ persons} \times 1 \text{ year}} = 0.008\,104 \text{ year}^{-1}$ or 810.4 per 100 000 person-years.

In 2007, there were 2 423 712 deaths in the United States. The mid-year population size was estimated at 301 621 000. Thus, the mortality rate in 2007 was $\frac{2\,423\,712}{301\,621\,000 \text{ persons} \times 1 \text{ year}} = 0.008\,036 \text{ year}^{-1} = 803.6$ per 100 000 person-years.

We need not limit ourselves to one-year periods of observation. For example, the two years 2006 and 2007 had an average population of (299 398 000 + 301 621 000)/2 = 300 509 500. There were (2 426 264 + 2 423 712) = 4 849 976 deaths over these two years. Therefore, the mortality rate for 2006 and 2007 combined is $\frac{\text{no. of incident events}}{(\text{average population size}) \times (\text{time of observation})} = \frac{4\,849\,976}{300\,509\,500 \text{ persons} \times 2 \text{ years}} = 0.008\,070 = 807.0$ per 100 000 person-years.

The concept of a rate can be flexibly applied to a variety of other ''risk-units,'' as demonstrated in Illustrative Example 3.6.

Illustrative Example 3.6 Traffic accident fatality rate

In 1994 there were 40 100 automobile-related fatalities and 2297 billion passenger-miles traveled in automobiles. Thus, the fatality rate associated with automobile travel in 1994 was $\frac{40\,100 \text{ fatalities}}{2297 \times 10^9 \text{ miles}} = 17.5$ fatalities per billion miles traveled.

Prevalence

Prevalence (**prevalence proportion, point prevalence)** refers to the proportion of individuals in a population that have a disease or condition at a specific point in time:

$$\text{Prevalence} = \frac{\text{no. of individuals with the condition}}{\text{no. of individuals considered}} \quad (3.5)$$

The numerator of a prevalence calculation includes all individuals with the condition under consideration regardless of when the disease commenced. The denominator is the total number of individuals under consideration. Prevalence can be calculated in both opened and closed populations.

Interpretation: Prevalence simply refers to the proportion of individuals in the population currently with a disease or condition. When based on a simple random sample from a population, the prevalence is an estimate of the probability that an individual currently has the condition in question.

Illustrative Example 3.7 Prevalence

A simple random sample of 1000 individuals from a population demonstrates 52 diabetics and 948 non-diabetics. Therefore, the prevalence of diabetes is $\frac{52 \cancel{\text{ persons}}}{(52+948) \cancel{\text{ persons}}} = 0.052$ or 5.2%.

Some epidemiologic sources consider a form of prevalence know as the **period prevalence**, which is

$$\text{Period prevalence} = \frac{\text{no. that experienced the condition during an interval}}{\text{total no. of individuals considered during the interval}} \quad (3.6)$$

Illustrative Example 3.8 Period prevalence

During the course of a semester, 23 of the 58 students in a class experienced at least one upper respiratory infection. Thus, the period prevalence of upper respiratory infections was $\frac{23}{58}$ or 39.7%.

Note: Period prevalences reflect some characteristics of incidence and some of prevalence. Therefore, some authorities recommend that we avoid use of the period prevalence and report separate incidence and point prevalence estimates instead (Elandt-Johnson and Johnson, 1980 p. 32).

The prevalence of a disease in a population depends on the rate of inflow of cases into the population and outflow of cases from the population. *Inflow* is determined by the incidence rate of the disease in the population and the immigration into the population of people who already have the disease. *Outflow* is determined by the rate of resolution either through recovery or death, and also by the emigration of cases from the population.

To understand the dynamics of prevalence, imagine the fluid level in a basin in which water flows in through incidence and drains out through death (Figure 3.6). The level of water in the basin represents the prevalence of the condition. Note that the prevalence of a condition can increase from either an elevation in incidence or decreases in the death rate. For example, improved survival of HIV/AIDS patients through effective treatment will increase the prevalence of the condition in the population if the incidence of HIV/AIDS remains constant.

Thus, the prevalence of disease is related to the duration of the disease according to this formula: prevalence $\approx$ (incidence rate) $\times$ (average duration of disease).[e] For

[e] The formula assumes the population is stationary, the disease is rare, and the incidence of a disease and its duration are constant over time. It is presented to clarify the conceptual relationship between incidence and prevalence, but it is rarely used in practice because of its complex population dynamic assumptions. Ahlo (1992) presents a more complex formula that can be used to derive the overall prevalence of a condition in a population based on age-specific incidence rates and average durations of disease.

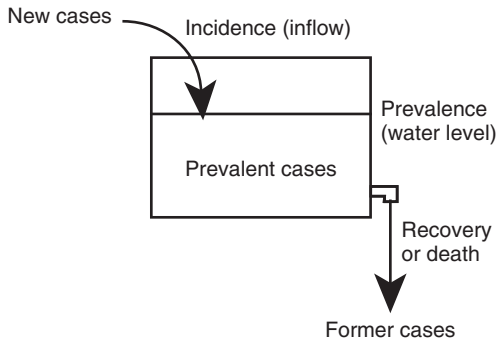


Figure 3.6 Dynamics of prevalence.

example, a disease with an incidence rate of 0.01 year^{-1} and average duration of $^1/_2$ year under steady-state conditions has prevalence $\approx$ 0.01 year^{-1} $\times$ 0.5 year = 0.005.

Comparison of incidence and prevalence: Incidence and prevalence represent distinct measures of disease frequency. Incidence addresses the transition from the disease-free state to the diseased state. In contrast, prevalence addresses current health. Thus, because it is linked to the duration of illness, prevalence is not as well suited as incidence for studying causation. Other differences between incidence and prevalence are summarized in Table 3.1.

3.2 Measures of association

Background

Measures of association in epidemiology are used to quantify the effect of an exposure on an outcome. Therefore, measures of association are also called **measures of effect**.

In measuring association, we will use the term **exposure** to denote any explanatory factor thought to increase or decrease the likelihood of the health outcome under consideration. We will also use the term **disease** to denote any dependent variable or health outcome. For example, we may speak of (a) smoking as an exposure that causes lung cancer, (b) advanced maternal age at pregnancy as an exposure that causes Downs syndrome, (c) high dietary fat as an exposure that causes coronary artery disease, and

Table 3.1 Comparison of incidence and prevalence.

Incidence	Prevalence
Counts onsets of events only	Counts both ''new'' and ''old'' cases
Independent of mean duration of disease	Depends on mean duration of disease
Can be measured as a rate or proportion	Always measured as a proportion
Reflects likelihood of *developing* disease over time	Reflects likelihood of *having* disease at point in time
Preferred measure when studying disease etiology	Preferred measure when studying health services utilization

(d) improved fitness as an exposure that reduces overall mortality. ''Exposure'' and ''disease'' are jargon for ''explanatory variable'' and ''response variable,'' respectively.

Absolute versus relative comparisons

Measures of association are made by *comparing* the rate or risk of disease in an exposed group to that of a nonexposed group. Before addressing epidemiologic comparisons, let us review relevant arithmetic principles by comparing the weight of a man who weighs 100 kg with the weight of a woman who weighs 50 kg.

1. We may say the man is 50 kg heavier than the women. This is an **absolute comparison** since it is made in unqualified terms. Note that absolute comparisons are made by **subtraction** and hence retain the initial unit of measure—the man weighs 50 *kilograms* more than the women.
2. Alternatively, we may say the man is twice as heavy as the woman. This is a **relative comparison** since it is made relative to the woman's weight. Relative measures of comparison are made by **division** and in the process made unit-free—the man is ''twice the weight'' of the woman.[f]

Let us now apply these arithmetic principles to measures of disease frequency. Say, an exposed group demonstrates a rate of 2 per 100 person-years, while a nonexposed group demonstrates a rate of 1 per 100 person-years.

1. In absolute terms, the exposure increased the rate by (2 case per 100 person-years) − (1 case per 100 person-years) = 1 case per 100 person-years. This **rate difference** represents the effect of the exposure in **absolute terms**.
2. Alternatively, we may say that the exposed group has $\frac{2 \text{ per } 100 \text{ person-years}}{1 \text{ per } 100 \text{ person-years}} = 2$ times (twice) the rate of the nonexposed. This **rate ratio** represents the effect of the exposure in **relative terms**.

Absolute measures of effect

As noted, absolute comparisons are made by subtraction. Thus, the **rate or risk difference (*RD*)** quantifies the effect of an exposure in absolute terms according to this formula:

$$RD = R_1 - R_0 \tag{3.7}$$

where R_1 represents the risk or rate of disease in the exposed group and R_0 represents the risk or rate in the nonexposed group. This formula may also be applied to prevalence ''rates,''[g] in which case it describes a **prevalence difference**.

Positive *RD*s indicate the excess rate associated with exposure in absolute terms. Negative *RD*s indicate the deficit in the rate or risk.

---

[f] We could also say the man weighs 100% *more* than the woman. However, it is incorrect to say that the man weighs 200% more than the woman, as this would imply that the man is three times as heavy.

[g] The term rate is included in parentheses to indicate the common misperception of prevalence as a rate. Mathematically, prevalence is a proportion and is not a rate.

Illustrative Example 3.9 Smoking and lung cancer (rate difference)

An important historical study found an age-adjusted lung cancer mortality of 104 per 100 000 person-years in doctors who smoked (Doll and Peto, 1976). Doctors who had never smoked had an age-adjusted lung cancer mortality rate of 10 per 100 000 person-years. Therefore, the $RD = R_1 - R_0 =$ (104 per 100 000 person-years) – (10 per 100 000 person-years) = 94 per 100 000 person-years. Thus, the effect of smoking was to increase lung cancer mortality by producing an additional 94 lung cancer deaths per 100 000 person-years.

Illustrative Example 3.9 demonstrated a positive association between the exposure and disease. Here is an example of a negative association.

Illustrative Example 3.10 Improved physical fitness and mortality (rate difference)

A study of physical fitness and overall mortality found that men who improved their physical fitness from the unfit level to the fit level had an age-adjusted death rate of 67.7 per 10 000 person-years (Blair *et al*., 1995). Men who remained unfit had an age-adjusted death rate of 122.0 per 10 000 person-years. Thus, improved physical fitness was associated with an $RD = R_1 - R_0 =$ (67.7 per 10 000 person-years) – (122.0 per 10 000 person-years) = – 54.3 per 10 000 person-years. This indicates 54.3 fewer deaths per 10 000 person-years associated with improved fitness.

Relative measures of effect

The **rate or risk ratio (*RR*)** quantifies the effect of an exposure in relative terms:

$$RR = \frac{R_1}{R_0} \tag{3.8}$$

where R_1 once again represents the risk or rate in the exposed group and R_0 represents the risk or rate in the nonexposed group. Formally, the ratio of two incidence rates is a rate ratio and the ratio of two incidence proportions is a risk ratio. When this formula is applied to the ratio of two proportions, it results in a prevalence ratio. All of these ratio measures of effect are referred to as **relative risks**.

Interpretation: The *RR* quantifies the excess (*RR*s greater than 1) or deficit (*RR*s less than 1) in the rate or risk of disease associated with exposure in relative terms. It is literally the **risk multiplier** associated with exposure. For example, an *RR* of 2 indicates that the exposure doubles the rate or risk of disease, while an *RR* of $1/2$ indicates that the exposure cuts the rate or risk in half.

Thus, the *RR* indicates both the **direction** and **strength** of an observed association. *RR*s greater than 1 indicate a positive association; those less than 1 indicate a negative association. Just as importantly, the further the *RR* gets from 1, the stronger the association. For example, an *RR* of 3 indicates a stronger positive association than an *RR* of 2. Analogously, an *RR* of $1/3$ indicates a stronger negative association than an *RR* of $1/2$.

Illustrative Example 3.11 Intravenous drug use and HIV (prevalence ratio)

A seroprevalence survey performed in the New York State female prison population revealed that 61 of 136 (44.85%) intravenous drug users were HIV positive. In contrast, 27 of 339 (7.96%) of non-users were HIV positive (Smith *et al*., 1991). Therefore, $RR = \frac{R_1}{R_0} = \frac{44.85\%}{7.96\%} = 5.63$. This indicates that the

prevalence of HIV in the exposed (intravenous drug user) group was 5.6 times that of the nonexposed group.

Note: Prevalence ratios will be equivalent to risk ratio when the disease outcome is rare (risk less than or equal 5%), the mean duration of disease among the exposed and nonexposed cases is the same, and developing the disease does not change the exposure status of study subjects.

Here is an example of a negative association expressed as a rate ratio.

Illustrative Example 3.12 Improved physical fitness and mortality (rate ratio)

Recall the physical fitness and mortality data used in Illustrative example 3.10. The adjusted mortality rate in men who improved their fitness was 67.7 per 10 000 person-years. The mortality rate in those who did not improve their fitness was 122.0 per 10 000 person-years. Therefore, the rate ratio $= \frac{67.7 \text{ per } 10\,000 \text{ person-years}}{122.0 \text{ per } 10\,000 \text{ person-years}} = 0.55$. This negative association indicates that improved fitness was associated with cutting mortality *almost* in half.

Relative risk difference: The relative risk difference (*RRD*) is an alternative expression of the *RR* that is derived by subtracting 1 from the *RR*: $RRD = RR - 1$. This statistic expresses the risk difference relative to the baseline risk established by the nonexposed group,[h] and offers an effective way to explain relative risks to the public. For example, the rate ratio of 0.55 in Illustrative Example 3.12 can now be expressed as $RRD = (RR - 1) = 0.45 - 1 = -0.45$, indicating a 45% *reduction* in mortality with improved fitness. The prevalence ratio in Illustrative Example 3.13 can be expressed as $RRD = 5.63 - 1 = 4.63$, indicating 463% *greater* prevalence in the intravenous drug user group. This expression is more palatable than the alternative ''a prevalence that is 5.63 times that of the non-IV drug users.''

Odds ratios

The odds ratio (*OR*) provides an alternative measure of relative effect. However, instead of being based on proportions, it is based on odds.

The odds of an event is simply its ratio of ''successes'' to ''failures.'' For example, if 1 in 5 people experience an adverse event, the risk of the event is 1 in 5 (20%) but its odds are 1 to 4 (0.25). Odds may be used in place of incidence proportions (risk) and prevalences, but cannot be applied to person-time data where the number of ''failures'' (non-cases) is not available.

Let us use the notation in Table 3.2 to contemplate *OR*s. Using this notation, *A* represents ''case'' and *B* represents ''non-case,'' while the subscript ''$_1$'' represents ''exposed'' and subscript ''$_0$'' represents ''nonexposed.'' For example, A_1 represents the number of exposed cases and A_0 represents the number of nonexposed cases.

---

h

$$\text{Relative risk difference} = \frac{\text{risk difference}}{\text{baseline risk}} = \frac{R_1 - R_0}{R_0} = \frac{R_1}{R_0} - \frac{R_0}{R_0} = RR - 1.$$

Table 3.2 Notation for cross-tabulated counts.

	Disease +	Disease −	Total
Exposure +	A_1	B_1	N_1
Exposure −	A_0	B_0	N_0

Using this notation, the odds of disease in the exposed group is A_1/B_1 and the odds in the nonexposed group is A_0/B_0. The ratio of these odds is:

$$OR = \frac{A_1/B_1}{A_0/B_0} \tag{3.9}$$

or equivalently,

$$OR = \frac{A_1 B_0}{B_1 A_0} \tag{3.10}$$

Interpretation: The *OR* is most often interpreted as if it were an *RR*. This is because, when the disease outcome is rare, $OR \approx RR$. However, the *OR* is also an effective measure of association in its own right, expressing the relative odds of the outcome in the exposed and nonexposed groups.

Illustrative Example 3.13 Folic acid and neural tube defects (odds ratio)

Neural tube defects (e.g., spina bifida) are a common type of birth defect affecting approximately 4000 pregnancies annually in the United States (CDC, 1992). Milunsky *et al*. (1989) examined the relationship between the use of folic acid-containing vitamins around the time of conception and neural tube defects in an HMO population. All the study subjects were undergoing maternal screening. Ten of the 10 713 women who used multivitamins that contained folic acid during the first 6 weeks of pregnancy were reported to have had a baby with a neural tube defect. In comparison, 11 of 3157 pregnancies in women who had not used multivitamins before or after conception had a baby with a neural tube defect. Data are shown in Table 3.3. The $OR = \frac{A_1 B_0}{B_1 A_0} = \frac{10 \cdot 3146}{10703 \cdot 11} = 0.27$ indicating a 73% reduction in neural tube defects associated with folic acid containing multivitamins.

Relation between the *RR* and *RD*

The risk ratio (*RR*) and risk difference (*RD*) describe different aspects of the association between an exposure and disease. As noted earlier, *RR*s provide relative measures of effect, while *RD*s provide absolute measures of effect.

Table 3.3 Data for Illustrative Example 3.13, folic acid and neural tube defects (NTD).

	NTD +	NTD −	Total
Folic acid +	10	10 703	10 713
Folic acid −	11	3146	3157
	21	13 849	13 870

Table 3.4 Adjusted mortality rates for lung cancer and ischemic heart disease in smokers and nonsmokers.

	Smokers (per 100 000 person-years)	Nonsmokers (per 100 000 person-years)	Rate difference (per 100 000 person-years)	Rate ratio
Lung cancer	104	10	94	10.4
Coronary heart disease	565	413	152	1.4

Source: Doll and Peto (1976).

Mathematically we note: $RD = R_1 - R_0$. Dividing both sides of this equation by R_0 derives $\frac{RD}{R_0} = \frac{R_1}{R_0} - \frac{R_0}{R_0} = RR - 1$. Since $\frac{RD}{R_0} = RR - 1$, then $RD = (RR - 1)R_0$. Scrutiny of this last expression reveals that the *RD* is the product of the segment of *RR* above 1 and the rate in the nonexposed group (R_0). Thus, even a large *RR* can have a modest *RD* when the disease is rare. In contrast, a small *RR* can have a large *RD* when the disease is common.

To see how this plays out in an epidemiologic context, consider the data in Table 3.4. Although the *RR* for smoking and lung cancer is much greater than the *RR* for smoking and heart disease (10.4 versus 1.4), the *RD* for smoking and lung cancer is far smaller than the *RD* for smoking and heart disease (94 per 100 000 versus 152 per 100 000). This is because heart disease is much more common than lung cancer. Therefore, the modest *RR* of 1.4 for heart disease greatly increases the number of cases in a population. In contrast, because lung cancer is relatively rare, the large *RR* associated with smoking translates to fewer additional cases.

3.3 Measures of potential impact

Attributable fraction in the population

The **attributable fraction in the population (AF_p)** is the difference between the current population rate and the rate associated with absence of the risk factor expressed as a fraction of the current population rate. Thus,

$$AF_p = \frac{R - R_0}{R} \tag{3.11}$$

where R (no subscript) represents the rate in the population as a whole and R_0 represents the rate in the absence of exposure. This statistic answers the question: ''What fraction of the disease burden in the population would potentially be averted with blanket removal of the exposure from the population?''

For example, the rate of lung cancer in the population as a whole (R) is approximately 15.9 per 100 000 person-years, while the rate in nonsmokers (R_0) is 3.5 per 100 000 person-years. Therefore, $AF_p = \frac{R - R_0}{R} = \frac{15.9 - 3.5}{15.9} = 0.78$, indicating that up to 78% of the lung cancer cases in this population are potentially preventable through the elimination of smoking.

An alternative (equivalent) formula for the AF_p is

$$AF_p = \frac{(p_e)(RR - 1)}{1 + (p_e)(RR - 1)} \tag{3.12}$$

where RR is the risk ratio associated with the exposure and p_e is prevalence of exposure in the population. For example, if 40% of the population smoked (p_e) and the RR of lung cancer associated with smoking in this population was 10, then

$$AF_p = \frac{(p_e)(RR - 1)}{1 + (p_e)(RR - 1)} = \frac{0.40 \cdot (10 - 1)}{1 + 0.40 \cdot (10 - 1)} = 0.78$$

Formula (3.12) demonstrates that the AF_p is a function of the strength of the association expressed as an RR and prevalence of the exposure, p_e.

Yet another alternative formula for the population attributable fraction is

$$AF_p = \frac{p_c(RR - 1)}{RR} \tag{3.13}$$

where p_c represents the proportion of *cases* in the population that are classified as exposed and RR represents the risk ratio. This formula is useful when working with case–control data (Chapter 8) in which the OR can substitute for the RR and p_c can be determined directly from the case series. Suppose, for example, that 87% of lung cancer cases in a case–control study smoked (p_c) and the odds ratio for smoking and lung cancer in this study is 10. Thus, $AF_p = \frac{p_c(RR - 1)}{RR} = \frac{0.87 \cdot (10 - 1)}{10} = 0.78$.

Table 3.5 lists estimates for population attributable fractions for various risk factors and cancer (all forms combined). Although these are only rough estimates, they are useful for indicating where preventive efforts should be focused to achieve the greatest potential reductions in cancer-related incidence and death. Note that interventions directed toward tobacco and diet have the greatest potential impact.

Table 3.5 Population attributable fractions for various types of modifiable risk factors and cancer incidence and mortality.

Risk factor type	*Attributable fraction*$_{population}$ (%)
Tobacco	29–30
Dietary	20–35
Occupational	4–9
Reproductive and sexual	7
Sunlight and background radiation	3–10
Pollution	2
Drugs and medical radiation	1–2
Industrial and consumer products	<1
Infective processes	5–10

Based on estimates published in Doll and Peto (1981), Miller (1992), Farrow and Thomas (1998), and Brownson *et al*. (1993).

Table 3.6 Range of population attributable fractions associated with modifiable risk factors for lung cancer, United States.

Risk factor	*Attributable fraction*$_{population}$
Cigarette smoking	80–90
Occupational exposures	10–20
Residential radon	7–25
Low vegetable diet	0–5
Environmental tobacco	0–2

Based on information in Farrow and Thomas (1998), Brownson *et al*. (1993), Reynolds *et al*. (1991), and Alberg *et al*. (2007).

Table 3.6 lists AF_ps for lung cancer and selected modifiable risk factors. Notice that the sum of the attributable fractions exceeds 100%. This should come as no surprise since removal of one component cause in a sufficient causal mechanism will prevent disease occurrence (see Section 2.3). Thus, any given case can be prevented in multiple ways.

Attributable fraction in exposed cases

The **attributable fraction in exposed cases (AF_e)** is:

$$AF_e = \frac{R_1 - R_0}{R_1} \tag{3.14}$$

where R_1 is the rate in the exposed population and R_0 is the rate in the nonexposed population. This statistic answers the question: ''What fraction of the exposed cases would have been averted if they had not been exposed to the risk factor in question?''

Algebraic manipulation of Formula (3.14) derives this equivalent formula:

$$AF_e = \frac{RR - 1}{RR} \tag{3.15}$$

For example, the *RR* of lung cancer associated with moderate smoking in the United States has been estimated to be 10. Therefore, $AF_e = \frac{RR - 1}{RR} = \frac{10 - 1}{10} = 0.90$, suggesting that 90% of the lung cancer cases among moderate smokers could have been averted had they not smoked.

Relation between the AF_p and AF_e: The AF_e is the proportion of exposed cases attributable to the risk factor in question. Since no case can be attributed to exposure unless they are exposed, the proportion of cases in the population attributable to the exposure (AF_p) is equal to the product of AF_e and the proportion of population cases that are exposed to the risk factor in question (p_c):

$$AF_p = AF_e \times p_c \tag{3.16}$$

Suppose for example that a risk factor with an AF_e of 0.5 is present in 40% of the cases. Therefore, $AF_p = AF_e \times p_c = 0.5 \times 40\% = 20\%$.

Preventable fraction

The formulas for AF_e and AF_p do not allow for the calculation of attributable fractions associated with factors that decrease risk. One way to address this limitation is to interchange the definition of ''exposure'' in the study so that the group denied the beneficial factor is now denoted as ''exposed.'' This will result in an *RR* greater than 1, permitting application of the prior formulas.

Alternatively, we may directly calculate the *preventable fractions*. There are two types of preventable fractions. The *preventable fraction in the unexposed* is analogous to the attributable fraction in the exposed, and the *preventable fraction in the population* is analogous to the attributable fraction in the population

The **preventable fraction in the unexposed (PF_u)** is defined as $\frac{R_0 - R_1}{R_0}$ and is easily calculated with this algebraically equivalent formula:

$$PF_u = 1 - RR \tag{3.17}$$

This statistic answers the question: ''What proportion of unexposed cases could conceivably be prevented if exposed to the beneficial factor in question?'' This is synonymous with the **efficacy** of the intervention.

The **preventable fraction in the population (PF_p)** is defined as $\frac{R - R_1}{R}$, where R represents the rate in the population as a whole and R_1 represents the rate if everyone had been exposed to the beneficial factor. This statistic answers the question: ''What proportion of the disease in the population would be averted if the entire population were exposed to the beneficial factor?'' An equivalent formula for the preventable fraction in the population is:

$$PF_p = PF_u \times p_{cu} \tag{3.18}$$

where $PF_u = 1 - RR$ (Formula (3.17)) and p_{cu} represents the proportion of cases in the population that are unexposed to the beneficial factor in question.

Illustrative Example 3.14 Folic acid and neural tube defects (preventable fractions)

Recall Illustrative Example 3.13 in which we demonstrated that exposure to folic acid containing multivitamins early in pregnancy was effective in preventing neural tube defects. Table 3.3 lists the data for this illustration. Note that $RR = \frac{R_1}{R_0} = \frac{10/10713}{11/3157} = 0.27$. Therefore, the $PF_u = 1 - RR = 1 - 0.27 = 0.73$, indicating that 73% of the neural tube defects are potentially preventable with folic acid supplementation.

Also note that 11 of the 21 cases were not exposed to folic acid containing vitamins during the early stages of pregnancy. Therefore, $p_{cu} = 11/21 = 0.528$, and $PF_p = PF_u \times p_{cu} = 0.73 \times 0.528 = 0.39$, indicating that 39% of the cases in this population could potentially be averted if the population at risk was fully blanketed with folic acid supplementation.

3.4 Rate adjustment

In the first section of this chapter we measured disease frequency for an entire population. In this section we divide the population into subgroups based on age

to derive age-specific rates. We then present two methods for combining these age-specific rates to derive a single rate for the population that has been adjusted (''standardized'') for age. Because methods in this chapter apply to prevalences, incidence proportions, and incidence rates, we will once again use the term *rate* to refer to all three. This will allow us to present formulas and common ideas without redundancy.

Background

Age influences the rate of most diseases. Comparisons of rates must therefore account for age differences in populations. Otherwise, observed differences could be **confounded** by age and comparisons will be **biased**. Note that the term *bias* in epidemiology implies a systematic error in inference, *not* an imputation of prejudice due to partisanship or other factors. See Table 3.7 for several other important definitions we will use in this section.

A rate that applies to an entire population is called a **crude rate**. The problem of relying on crude rates is exemplified by the data presented in Table 3.8. This table contains rates for two different populations stratified by age. The stratification process divides a population into subgroups based on demographic factors such as age. Each subgroup is now called a **stratum**. In this illustration, we have two age strata. Note that the crude rate in population B is nine times that of the crude rate in population A (991 per 100 000 versus 109 per 100 000, respectively). Yet, when one compares rates within age-strata, the rates are identical. The explanation for this finding is that the disease is strongly age-related and population B is much older than

Table 3.7 Definitions of selected terms.

Bias: any process that can lead to a systematic error in inference
Confounding: the type of bias that comes about because of the effects of extraneous factors
Crude rate: a rate for the entire population
Stratum: a subgroup in a population defined by a specified criterion such as age, sex, or race
Stratum-specific rate: a rate for a population subgroup
Adjusted rate: a rate that has been mathematically or statistically manipulated so that the effects of differences in composition of the populations being compared have been minimized
Study population: the population for which rates are being adjusted
Reference population: an external population that provides an age (or other) distribution for adjustment procedures
Standard million: an age distribution for a million citizens selected at random from the reference population

Table 3.8 Rates in two populations stratified by age.

	Population A			Population B		
Age (years)	Cases	Person-years	Incidence rate (per 100 000 person-years)	Cases	Person-years	Incidence rate (per 100 000 person-years)
0–34	99	99 000	100	1	1000	100
35+	10	1000	1000	990	99 000	1000
Crude	109	100 000	109	991	100 000	991

Table 3.9 Standard million age distribution for the year 1991, United States.

Age group	Standard million
0–4	76 158
5–24	286 501
25–44	325 971
45–64	185 402
65–74	72 494
75+	53 474
Total	1 000 000

Source: U.S. Bureau of the Census (1994, p. 14).

population A. Thus, the comparison of crude rates was confounded by age differences. Age adjustment techniques (age standardization) will compensate for age differences like this and permit head-to-head comparisons of population rates.

Age adjustment (**age standardization**) can be achieved in several ways. The two most common approaches are by direct weighting and indirect weighting of strata-specific rates. Let us start with direct age-adjustment.

Direct adjustment

Direct age adjustment is a statistical procedure used to alleviate the effects of differences in age when comparing populations. This method uses the age distribution of an external **reference population** as the basis for comparison. The specific reference population used for this purpose is somewhat arbitrary but must be consistent within a given set of comparisons. A **standard million** is often used for this purpose. A standard million represents the age distribution for a million citizens picked at random from a population. Table 3.9 lists a standard million for the US population in 1991.

Direct adjustment creates a weighted average of the strata-specific rates from the **study population** (the population for which rates are being adjusted) based on the age distribution of the reference population. This derives the (hypothetical) rate of disease that would be expected in the reference population *if* it were to experience the age-specific rates of the study population. The formula for direct age adjustment is

$$\mathrm{aR}_{\mathrm{direct}} = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i} \tag{3.19}$$

where $\mathrm{aR}_{\mathrm{direct}}$ represents the adjusted rate by the direct method, subscript i is a stratum counter (there are k strata), N_i represents the number of people in stratum i of the reference population, r_i represents the rate of disease in stratum i of the study population, and $\sum$ denotes the summation operator defined as $\sum_{i=1}^{k} N_i = N_1 + N_2 +$

$\ldots + N_k$. Note that capital letters in this formula (N_i) denote values that come from the reference population, while lowercase letters (r_i) denote values from the study population. Also, be aware that multiplication takes precedence over addition in the order of mathematical operations, so that $\sum_{i=1}^{k} N_i r_i = (N_1 r_1) + (N_2 r_2) + \ldots + (N_k r_k)$.

Let us consider death rates in Alaska and Florida to illustrate use of this formula. Data are presented in Table 3.10. Note that the crude death rate (cR) per 100 000 Alaska residents is 387, while the crude death rate per 100 000 Floridians is 1026. Based on these crude rates it might appear that Florida is a riskier environment than Alaska. However, given Florida's proclivity for attracting retirees, we might ask if the observed difference in crude rates is due to age differences in the two populations. To address this question we will use the direct method of rate adjustment on both populations.

The first step in adjusting for age by the direct method is to calculate age-specific rates in both populations. These values are shown in Table 3.11. From these data, we note that age-specific death rates vary only slightly among the states.

Let us use the 1991 standard million US population (Table 3.9) as our reference age distribution. Thus, the age-adjusted mortality rate in Alaska is 843 per 100 000 (see Table 3.12 for calculations); the age-adjusted mortality rate in Florida is 784 per 100 000 (see Table 3.13). Thus, the initial excess in Florida disappears after age-adjustment. Florida has a lower mortality rate after adjusting for age.

Indirect adjustment

The indirect rate method of adjustment ($aR_{indirect}$) is based on multiplying the crude rate (cR) in the reference population by a ratio known as the standardized mortality/morbidity ratio (SMR):

$$aR_{indirect} = (cR)(SMR) \tag{3.20}$$

Table 3.10 Vital statistics for Alaska and Florida, 1991.

	Alaska		Florida	
Age (years)	Deaths[a]	Population[b]	Deaths[c]	Population[d]
0–4	122	57 000	2177	915 000
5–24	144	179 000	2113	3 285,000
25–44	382	222 000	8400	4 036 000
45–64	563	88 000	21 108	2 609 000
65–74	406	16 000	30 977	1 395 000
75+	582	7000	71 483	1 038 000
Totals	2200	569 000[a]	136 258	13 278 000[a]

[a] Source: NCHS (1993, p. 101).
[b] Source: U.S. Bureau of the Census (1992, p. 26).
[c] Source: NCHS (1993, p. 105). Age not stated for 35 decedents (omitted from table).
[d] Total may not sum accurately due to rounding error.
cR_{Alaska} (per 100 000 residents) = 2200/569 000 × 100 000 = 387.
$cR_{Florida}$ (per 100 000 residents) = 136 258/13 278 000 × 100 000 = 1026.

Table 3.11 Age-specific death rates per 100 000 residents for Alaska and Florida, 1991. Based on data in Table 3.10

Stratum	Age (years)	Alaska	Florida
1	0–4	214[a]	238
2	5–24	80	64
3	25–44	172	208
4	45–64	640	809
5	65–74	2538	2221
6	75+	8314	6887

[a] Example of calculation: age-specific rate, 0–4-year-olds, Alaska, per 100 000.

$$r_{\text{stratum 1, Alaska}} = \frac{\text{deaths}}{\text{population}} \times \text{multiplier} = \frac{122}{57\,000} \times 100\,000 = 214$$

Table 3.12 Calculation of the age-adjusted death rate for Alaska, 1991, using the standard million from Table 3.9 as the external reference population.

Stratum (i)	Age (years)	Rate (per 100 000) (r_i)	Standard million (N_i)	Product ($N_i r_i$)
1	0–4	214	76 158	16 297 812
2	5–24	80	286 501	22 920 080
3	25–44	172	325 971	56 067 012
4	45–64	640	185 402	118 657 280
5	65–74	2 538	72 494	183 989 772
6	75+	8 314	53 474	444 582 836
		Column sums →	1 000 000	842 514 792

Calculations:

$$\sum N_i r_i = N_1 r_1 + N_2 r_2 + \ldots + N_6 r_6 = 16\,297\,812 + 22\,920\,080 + \ldots + 444\,582\,836 = 842\,514\,792$$

$$\sum N_i = N_1 + N_2 + \ldots + N_6 = 76\,158 + 286\,501 + \ldots + 53\,474 = 1\,000\,000$$

$$\text{aR}_{\text{direct}} = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i} = \frac{842\,514\,792}{1\,000\,000} \approx 843$$

The SMR is the ratio of the observed number of cases in a population (A) to the expected number (μ):

$$\text{SMR} = \frac{A}{\mu} \tag{3.21}$$

The expected number of cases, μ, is calculated according to Formula (3.22):

$$\mu = \sum_{i=1}^{k} R_i n_i \tag{3.22}$$

Table 3.13 Calculation of the age-adjusted death rate for Florida, using the standard million from Table 3.9 as the external reference population.

Stratum (i)	Age (years)	Rate (per 100 000) (r_i)	Standard million (N_i)	Product ($N_i r_i$)
1	0–4	238	76 158	18 125 604
2	5–24	64	286 501	18 336 064
3	25–44	208	325 971	67 801 968
4	45–64	809	185 402	149 990 218
5	65–74	2221	72 494	161 009 174
6	75+	6887	53 474	368 275 438
		Column sums →	1 000 000	783 538 466

Calculations:

$$\sum N_i r_i = N_1 r_1 + N_2 r_2 + \ldots + N_6 r_6 = 18\,125\,604 + 18\,336\,064 + \ldots + 368\,275\,438 = 783\,538\,466$$

$$\sum N_i = N_1 + N_2 + \ldots + N_6 = 76\,158 + 286\,501 + \ldots + 53\,474 = 1\,000\,000$$

$$aR_{direct} = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i} = \frac{783\,538\,466}{1\,000\,000} \approx 784$$

where R_i represents the rate of disease in the ith stratum of the reference population and n_i represents the number of people in the ith stratum of the study population. The product $R_i n_i$ is the expected number of cases in the ith stratum of the study population (μ_i), assuming the study population has the same underlying risk as the comparable stratum in the reference population. Note once again, that capital letters (R_i) denote values that come from the reference population and lowercase letters (n_i) denote values from the study population.

To illustrate age adjustment by the indirect method, we compare death rates in Zimbabwe and the United States. The crude death rate in Zimbabwe in the early 1990s was 886 per 100 000 (Table 3.14). The crude death rate in the United States in

Table 3.14 Vital statistics for Zimbabwe.

Age (years)	Deaths[a]	Population[b]	Rate (no multiplier)
0–4	?[c]	1 899 204	?
5–24	?	5 537 992	?
25–44	?	2 386 079	?
45–64	?	974 235	?
65–74	?	216 387	?
75 +	?	136 109	?
Total	98 808	11 150 006	0.008 86

[a] Date are for 1992. Source: United Nations (1996, p. 138).
[b] Date are for 1994. Source: United Nations (1996, pp. 186–187).
[c] ? = Not known.
$cR_{Zimbabwe} = 98\,808/11\,150\,006 = 0.008\,86 = 886$ per 100 000.

Table 3.15 Vital statistics for the United States, 1991

Age (years)	Deaths[a]	Population[b]	Rate (no multiplier)
0–4	44 000	19 204 000	0.002 29
5–24	45 000	72 244 000	0.000 62
25–44	147 700	82 197 000	0.001 80
45–64	368 800	46,751,000	0.007 89
65–74	478 600	18 280 000	0.026 18
75+	1 084 900	13 484 000	0.080 46
Total	2 169 000	252 160,000	0.008 60

[a]Source: NCHS (1993, pp. 101–102).
[b]Source: U.S. Bureau of the Census (1992, p. 26).
Reference rate (crude U.S. mortality rate = 2 169 000/252 160 000 = 0.008 60 = 860 per 100 000.

1991 was 860 per 100 000 (Table 3.14). On inspection, we determine that Zimbabwe's population is much younger than the United States' (Figure 3.7). Given the observed difference in age distributions, we might ask what the difference in death rates is after age has been accounted for.

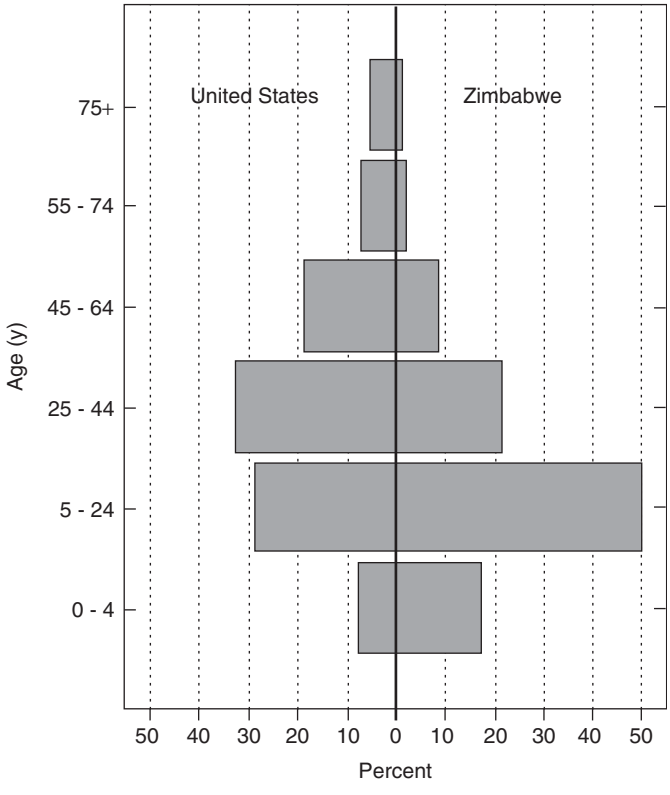


Figure 3.7 Population pyramids for the United States and Zimbabwe, 1991 and 1994, respectively. (Data sources: U.S. Bureau of the Census, 1992; United Nations, 1996.)

Table 3.16 Calculation of the adjusted rate by the indirect method for Zimbabwe, using the United States, 1991, as the external reference population.

Stratum (i)	Age (years)	US rate (R_i)	Zimbabwe population (n_i)	Product ($R_i n_i$)
1	0–4	0.002 29	1 899 204	4349
2	5–24	0.000 62	5 537 992	3434
3	25–44	0.001 80	2 386 079	4295
4	45–64	0.007 89	974 235	7687
5	65–74	0.026 18	216 387	5665
6	75+	0.080 46	136 109	10 951
			Column sum (μ)→	36 381

$$cR_{Zimbabwe} = 98\,808/11\,150\,006 = 0.008\,86 = 886 \text{ per } 100\,000$$

$$\mu = \sum R_i n_i = R_1 n_1 + R_2 n_2 + \ldots + R_6 n_6 = 4349 + 3434 + \ldots + 10\,951 = 36\,381$$

$$SMR = \frac{A}{\mu} = \frac{98\,808}{36\,381} = 2.716$$

$$aR_{indirect} = (cR)(SMR) = (0.008\,60)(2.716) = 0.023\,36 = 2336 \text{ per } 100\,000$$

Age-specific death rates in Zimbabwe are unavailable, but the age distribution of the population is known (Table 3.14). (This is an advantage of the indirect method: the age distribution of cases in the study population need not be known.) To adjust the Zimbabwe death rate using the indirect method, we apply age-specific death rates from the United States to determine the number of expected deaths in Zimbabwe. This value is used to calculate the SMR and the indirectly adjusted death rate, as shown in Table 3.16. Note that the SMR is 2.72. This number is a relative measure of effect—a ''relative risk''—which indicates the direction and strength of the association (see Section 3.2).

Once the SMR is calculated, it can be used to derive the indirectly adjusted death rate according to Formula (3.20). Thus, the indirectly adjusted rate for Zimbabwe is equal to (cR)(SMR) = (0.008 60)(2.716) = 0.023 36 or 2336 per 100 000 person-years. This age-adjusted rate can now be compared directly to the US death rate of 860 per 100 000.

Adjustment for multiple factors

The techniques in Section 3.4 have been presented in the context of adjustment for age differences among populations. However, these same techniques may also be applied to factors other than age, and can be used to adjust for multiple factors simultaneously. For example, data can be stratified by age (6 age groups), sex (2 groups), and race (3 groups) to derive 6 × 2 × 3 = 36 age-, sex-, race-specific rates (e.g., 0–4 year-old African American males). Direct- and indirect-adjustment methods directed toward these stratified data will then adjust for these multiple factors. Rates from different populations can then be compared with less concern[i] for confounding

[i] There may still be residual confounding due to the broad categories used to classify factors and due to misclassification.

by these factors. Direct and indirect adjustment techniques provides a flexible method to control for the confounding effects of many types of extraneous factors.

Section summary

1 Age adjustment (age standardization) is used to eliminate or reduce the confounding effects of age when comparing rates in populations that have different age distributions. Before rates can be adjusted, data must be stratified by age. Strata-specific information is then used to derive a single adjusted rate using either direct or indirect adjustment methods.
2 Direct age adjustment applies age-specific rates from the study population to an age distribution from a reference population. This provides the expected rate of disease in the reference population if it were to experience the same age-specific rates as the study population. Rates adjusted in this manner can then be compared with less concern for confounding by age.
3 Indirect age adjustment uses the age-specific rates from an external reference population to derive the expected number of cases in the study population. The expected number of cases is used to calculate a useful statistic known as the standardized mortality ratio (SMR), which can then either be interpreted directly or can be used to adjust the rate indirectly.

Notation used in Section 3.4

A	Observed number of cases
aR_{direct}	Adjusted rate, direct method (Formula 3.19)
$aR_{indirect}$	Adjusted rate, indirect method (Formula 3.20)
SMR	Standardized mortality ratio (Formula 3.21)
μ	Expected number of cases (Formula 3.22)
cR	Crude (unadjusted) rate
i	Stratum counter (used as a subscript)
n_i	Number of people, stratum i, study population
N_i	Number of people, stratum i, reference population
r_i	Rate, stratum i, study population
R_i	Rate, stratum i, reference population

Exercises

3.1 Point prevalences, period prevalence, and risk. The schematic shown as Figure 3.8 depicts a closed population of $N = 10$ individuals. Each individual, A through J, is followed from January 1 to December 31. The thick area of each line represents a period of illness, and the thin area represents a period of wellness. No subjects are lost to follow-up, and the disease confers lifelong immunity upon recovery. Based on these assumptions:

(A) What is the point prevalence of disease on January 1?

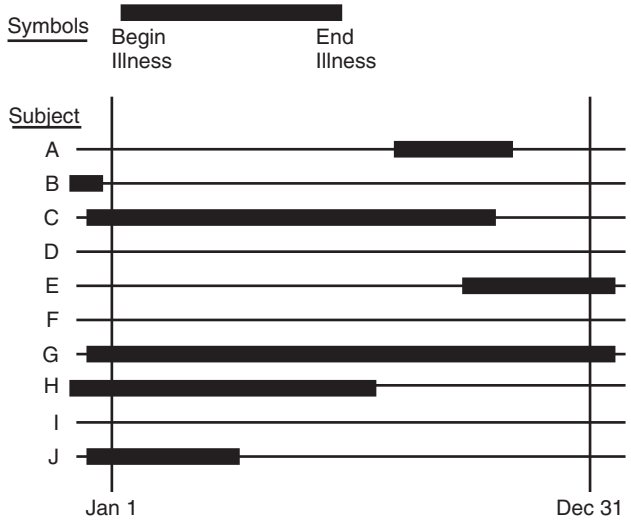


Figure 3.8 Schematic for Exercise 3.1.

(B) What is the point prevalence on December 31?
(C) What is the period prevalence for the interval January 1 to December 31?
(D) What is the incidence proportion (risk) for the year?

3.2 Hypertension in a cohort of men. A study of hypertension begins with one thousand 40- to 45-year-old men. Of these, 50 are already hypertensive. The remaining 950 are followed for 5 years, during which time 64 develop hypertension. (Assume no loss to follow-up or death due to competing risk.)
(A) Calculate the prevalence of hypertension at the beginning of the study.
(B) Calculate the 5-year incidence proportion (risk) of hypertension.
(C) Calculate the incidence rate of hypertension in the cohort with and without an actuarial adjustment. Did the actuarial adjustment make a difference? Explain why.

3.3 Vital statistics. Formulas for selected population based rates per m people are:

$$\text{Crude birth rate (per } m) = \frac{\text{no. of births}}{\text{midyear population}} \times m \tag{3.23}$$

$$\text{Crude death rate (per } m) = \frac{\text{no. of deaths}}{\text{midyear population}} \times m \tag{3.24}$$

$$\text{Infant mortality rate (per } m) = \frac{\text{no. of deaths} < 1 \text{ year of age}}{\text{no. of live births}} \times m \tag{3.25}$$

$$\text{Age-specific death rate (per } m) = \frac{\text{no. of deaths in age group}}{\text{midyear no. of people in age group}} \times m \tag{3.26}$$

$$\text{Cause-specific rate (per } m) = \frac{\text{no. deaths due to cause}}{\text{middyear total population}} \times m \tag{3.23}$$

Use these data to calculate each of statistics requested below.

Total midyear population	100 000
Population size, 65 years of age or older	25 000
Number of infants born alive	3000
Total deaths (all causes)	1500
Deaths of infants under 1 year of age	50
Deaths of persons 65 years of age and over	1000
Deaths from heart disease	300
Deaths from cancer	100

(A) The crude birth rate per 1000.
(B) The crude death rate per 1000.
(C) The infant mortality rate per 1000.
(D) The age-specific death rate for people 65 years of age or older, per 1000.
(E) The cause-specific death rate for heart disease, per 1000.
(F) The cause-specific death rate for cancer per 1000.

3.4 Prevalence in an open population. What effect will each of the following have on the prevalence of disease in a population assuming population dynamics and other elements of the dynamics of disease do not otherwise change?
(A) Immigration of cases into the population.
(B) Emigration of cases out of the population.
(C) Emigration of healthy persons out of the population.
(D) Immigration of healthy persons into the population.
(E) Increases in the case fatality rate.

3.5 Vital statistics for the United States, 1992. Base your answers for this question on the statistics in this table:

Population size	255 078 000
Approximate number of live births	4 065 014
Number of deaths (all ages)	2 175 631
Approximate number of deaths in infants under 1 year of age	34 553

Source: NCHS, 1995.

(A) Compute the birth rate per 1000.
(B) Compute the overall death rate per 100 000.
(C) Compute the infant mortality rate per 1000.

3.6 Effect of a treatment. Suppose a treatment is developed that prolongs life but does not result in a cure.
(A) How would this affect the incidence of the disease?

(B) How would this affect the prevalence of the disease?

3.7 Fatalities associated with travel. According to the *Statistical Abstract of the United States* (U.S. Bureau of the Census, 1995), in 1993, there were approximately 40,100 traffic fatalities. During this same period, there were approximately 2297 billion miles traveled in motor vehicles (i.e., cars, buses, and trucks).

(A) Calculate the rate of traffic fatalities per 100 million miles traveled.

(B) In 1993, the worldwide airline fatality rate associated with scheduled air transportation flights was 0.05 fatalities per 100 million miles flown. How does this compare with the rate of fatalities in motor vehicles?

3.8 Accidents in hospitals. The article on accidents in hospitals listed the age distribution of injuries in 82 injured patients as follows:

Age (years)	Number of accidents
0–2	5
3–5	6
6–14	18
15–21	8
22–31	5
32–41	8
42–51	7
52–61	4
62 and over	21

Based on these data, the authors concluded ''Statistical analysis by patient age group reveals that patients 62 years of age and older are most prone to accidents. The next greatest risk occurs among the 6– 14 year old age group.'' Comment on the authors' misinterpretation of the data.

3.9 Stationary? A stationary open population is one that maintains a constant size and age distribution. Over the 20th century, was the US population stationary? (Recall the discussion of the demographic transition presented in Section 1.2.)

3.10 Mortality rate and life expectancy. Figure 3.9 contains a two person-time drawing representing the survival experience of two cohorts on $n = 2$ each. Each line in this schematic represents an individual's years of life lived; D represents the point of death. Show that the average lifespan is equal to the reciprocal of the mortality rate within each population.

3.11 Comparing prevalences. The prevalence of a condition in population A is 20 per 100 000. The prevalence in population B is 10 per 100 000. The groups have identical age distributions. Can we conclude that population A has twice the incidence of population B of this condition? Explain.

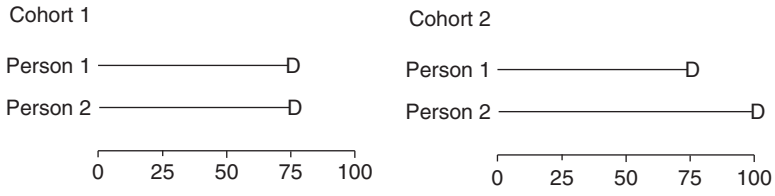


Figure 3.9 Schematic for Exercise 3.10.

3.12 Risk and rate of breast cancer. A cohort study starts with 10,000 women. Of these, 500 had experienced breast cancer sometime in their past. The remaining 9500 are followed for five-years. Two-hundred and fifty breast cancer cases occurred during the follow-up period. Assume there is no loss to follow-up and no competing risk in this cohort.

(A) What is the five-year incidence proportion (risk of breast cancer? Report the risk per 1000 people.

(B) What is the rate of breast cancer in this cohort? Report the rate per 1000 person-years.

(C) Show how risk $\approx$ rate $\times$ time using the results from (A) and (B) of this problem.

3.13 Coronary heart disease. One thousand people are approached to participate in a cohort study; 850 agree to participate; 50 have evidence of coronary heart disease upon their initial examination. Over the next 10 years, 100 of the disease-free study subjects develop coronary heart disease.

(A) What is the 10-year average risk of coronary heart disease in this cohort?

(B) What is the rate of coronary heart disease in the cohort?

3.14 Driving errors. ''Every 2 miles, the average driver makes 400 observations, 40 decisions, and one mistake. Once every 500 miles, one of those mistakes leads to a near collision, and once every 61,000 miles one of those mistakes leads to a crash'' (Gladwell, 2001).

(A) What is the rate of mistakes per mile?

(B) What is the risk an observation will be mistaken?

(C) What are the odds of near collisions to crashes?

3.15 $N = 6$. Figure 3.10 is a schematic in which each line represents individual follow-up and ''D'' represents disease onset.

(A) Talley the person-time is in this population.

(B) Determine the average number of people in the population at any given time with this formula: average number of individuals = (sum of person-time)/(time observed).

(C) Calculate the incidence rate using Formula (3.4).

3.16 Cohort study. An epidemiologist recruits 150 people for a cohort study. Of these potential study subjects, 10 are prevalent cases. The remaining study subjects are followed for 5 years, during which time 16 individuals develop the disease being studied.

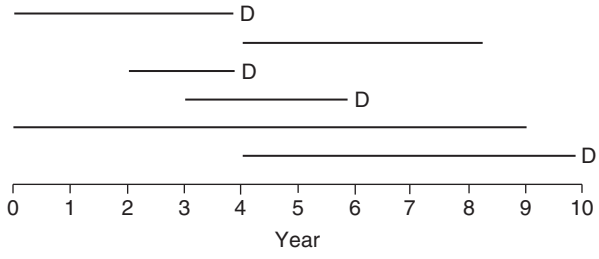


Figure 3.10 Person-time drawing for Exercise 3.15.

(A) What is the prevalence of disease at the start of the study?
(B) Assume all of the cases are still around at the end of the study. What is the prevalence now?
(C) What is the incidence proportion over the period of observation?
(D) What is the incidence rate over the period of observation?

3.17 More population based rates. A population demonstrates the following vital statistics:

Total midyear population	25 000
Population size, 65-years of age or older	2500
Number of live births	300
Total deaths (all cause)	250
Deaths in under 1-year olds	3
Deaths in persons 65 and over	75

(A) Calculate the birth rate per 1000.
(B) Calculate the mortality rate per 1000.
(C) Calculate the infant mortality rate per 1000.
(D) Calculate the mortality rate for those over 65 (per 1000).

3.18 Actuarial adjustment of person-time. A cohort of 100 individuals at risk is followed for 2 years. During that time, 18 individuals develop disease. We can say, roughly, that the cohort has 100 persons × 2 years = 200 person-years of observation time. However, the 18 incident cases do not contribute time at risk after developing disease. Therefore, the cohort has somewhat less than 200 person-years of observation time. How much person-time does the cohort have if we apply an actuarial adjustment?

3.19 Framingham men. The landmark Framingham Heart study was initiated in 1947 to study the epidemiology of heart disease in healthy volunteers. Since then, it has provided insights into risk factors for cardiovascular disease and stroke. This exercise considers data from one of its earliest publications (Kannel *et al.*, 1961). After 6 years of follow-up of men between the ages of 40 and 59, there were 16 coronary heart disease incidents among the 454 men with initial

cholesterol levels of less than 210 mg/100 ml (low serum cholesterol group). Among the 424 men with initial cholesterol levels of at least 245 mg/100 ml (high serum cholesterol group), there were 51 coronary heart disease incidents. Calculate the *RR* associated with high serum cholesterol and briefly, in plain language interpret this finding.

3.20 Framingham women. Let us continue to consider the 1961 report on cardiovascular incidence from the Framingham study. Of the 445 women in the 40–59 year old age group with initial cholesterol levels less than 210 mg/100 ml, there were 8 CHD onsets. Of the 689 women in this age group with cholesterol values of at least 245 mg/100 ml, there were 30 CHD onsets. Calculate the *RR* associated with high serum cholesterol in women and interpret your findings.

3.21 Restenosis. Each year cardiologists open clogged arteries only to have these same arteries undergo restenosis (recurrent narrowing) in about half of their patients. Until recently, no one has been able to accurately predict which patients will experience this adverse outcome. A study by Zhou and colleagues (1996) was conducted to determine whether there was an association between evidence of infection with cytomegalovirus and restenosis. After 6 months of follow-up, these researchers found restenosis in 21 out of 49 patients with evidence of prior cytomegalovirus infection. In comparison, 2 of the 26 patients lacking evidence of prior cytomegalovirus infection had a comparable degree of restenosis.

(A) Show these data in a 2-by-2 table. Then, calculate the risk ratio of restenosis associated with cytomegalovirus infection. Interpret this finding.

(B) Now calculate the odds ratio of restenosis. How does the odds ratio compare with the risk ratio calculated in part (A) of the exercise? Explain why there is a discrepancy.

3.22 Primary cardiac arrest and vigorous exercise. Siscovick and colleagues (1984) examined the rate of primary cardiac arrest associated with various levels of habitual high-intensity activity. Based on the data in this table, calculate the *RR*s associated with each level of habitual high-level activity using the 0 min/week group as the baseline for each comparison. Comment on your findings.

Habitual high-intensity activity (min/week)	Rate of primary cardiac arrest per 10^8 person-hours
0	18
1–19	14
20–139	6
≥140	5

Source: Siscovick *et al*. (1984).

3.23 California mortality. Table 3.17 reports vital statistics for the state of California in 1991.

(A) Calculate the crude death rate for the state. Compare this rate with that of Florida (Table 3.10).

Table 3.17 Vital statistics for California, 1991 (Exercise 3.23)

Age (years)	Deaths[a]	Population[b]
0–4	5500	2 651 000
5–24	5736	8 824 000
25–44	19 178	10 539 000
45–64	37 313	5 179,000
65–74	45 306	1 874 000
75 +	102 078	1 314 000
Total	215 111	30 381 000

[a]Source: NCHS (1993, p. 102).
[b]Source: U.S. Bureau of the Census (1992, p. 26).

(B) Calculate age-specific death rates.
(C) Using the standard million reported in Table 3.9 as the external reference population, directly adjust California's death rate.
(D) Compare California's adjusted death rate with that of Florida (Table 3.13).

3.24 Arkansas mortality. Table 3.18 reports vital statistics for the state of Arkansas in 1991.
(A) Calculate the crude death rate for the state. Compare this rate with that of Florida (Table 3.10).
(B) Calculate age-specific death rates.
(C) Using the standard million reported in Table 3.9 as the external reference population, directly adjust Arkansas's death rate.
(D) Compare Arkansas's adjusted death rate with that of Florida (Table 3.13).

3.25 Egyptian mortality. Table 3.19 reports vital statistics for Egypt.
(A) Calculate Egypt's crude death rate. How does this compare with the 1991 US crude death rate of 860 per 100 000.
(B) Adjust Egypt's death rate using the indirect method, using the US data reported in Table 3.15 as the external reference population. Interpret your results.

Table 3.18 Vital statistics for Arkansas, 1991 (Exercise 3.24).

Age (years)	Deaths[a]	Population[b]
0–4	449	170 000
5–24	562	697 000
25–44	1459	694 000
45–64	4072	458 000
65–74	5466	196 000
75 +	13 037	157 000
Total	25 048	2 219 000

[a]Source: NCHS (1993, p. 102).
[b]Source: U.S. Bureau of the Census (1992, p. 26).

Table 3.19 Vital statistics for Egypt (Exercise 3.25)

Age (years)	Deaths[a]	Population[b]
0–4	?[c]	7 909 000
5–24	?	24 560 000
25–44	?	13 764 000
45–64	?	6 921 000
65–74	?	1 485 000
75+	?	524 000
Total	416 000	55 163 000

[a]Data are for 1994. Source: United Nations (1996, p. 136).
[b]Data are for 1992. Source: United Nations (1996, pp. 178–179).
[c]?= Not known.

Review questions

R.3.1 What is the primary distinction between incidence and prevalence?

R.3.2 More people die each year in New York City than in Fairbanks, Alaska. Does this necessarily mean that New York City is a riskier place to live? Explain.

R.3.3 The number of cases in a population must be considered in relation to the _____ of the population that generated the cases.

R.3.4 What is a *ratio*?

R.3.5 Provide a synonym for *closed population*.

R.3.6 List synonyms for *incidence proportion*.

R.3.7 What goes into the numerator of an incidence proportion? What goes into the denominator?

R.3.8 Why do denominators of incidence proportions exclude those who are not at risk?

R.3.9 A group of women demonstrates a 5% risk of breast cancer. What additional information is needed to interpret this statement?

R.3.10 A group shows a one-year risk of 0.025. How many individuals are needed on average to generate one case?

R.3.11 List synonyms for *incidence rate*.

R.3.12 Propose three different ways to generate one *person-year*.

R.3.13 A carpenter works 60 hours fixing your kitchen. A tile layer works 8 hours on your kitchen. How many person-hours accumulated on the job?

R.3.14 What information goes into the numerator of an incidence rate? What goes into the denominator?

R.3.15 How many cases will a disease that has the rate of 0.013 33 person-year^{-1} generate in 1000 people observed for a year?

R.3.16 Under what condition is an incidence rate equal to an incidence proportion in a population?

R.3.17 What goes into the numerator of a prevalence calculation? What goes into the denominator of prevalence?

R.3.18 List ways in which prevalence differs from incidence.

R.3.19 If the rate of a disease remains constant, but the death rate due to the disease decreases over time through improved treatment, what happens to its prevalence of the condition over time?

R.3.20 What are the general terms that are often used to refer to the independent variable and dependent variable in epidemiologic studies?

R.3.21 True or false? The terms ''measure of association'' and ''measure of effect'' are used interchangeably.

R.3.22 In plain terms, what does it mean when we say there is a positive association between an exposure and disease?

R.3.23 What arithmetic operation is used to derive a relative comparison?

R.3.24 What arithmetic operation is used to derive an absolute comparison?

R.3.25 A report states that the exposure will cause an *additional* 5 cases per 1000 people. Is this an example of an *RR*, *RD*, *AF*e, or *AF*p?

R.3.26 A different report states ''the exposure doubles the risk.'' Is this an example of an *RR*, *RD*, *AF*e, or *AF*p?

R.3.27 What epidemiologic measure quantifies the effect of the exposure in *absolute* terms?

R.3.28 What epidemiologic measure quantifies the effect of an exposure in *relative* terms?

R.3.29 A report states ''people who do not wear seat belts are eight times as likely to die in an automobile crash than those how do.'' An *RR*, *RD*, *AF*e, or *AF*p?

R.3.30 The one-year risk in an exposed group is 15 per 1000. The one-year risk of disease in a nonexposed is 10 per 1000. Would it be correct say that the exposure increases risk by 150%? Explain your reasoning.

R.3.31 True or false? An *RR* of 0.7 indicates a positive association between the exposure and disease.

R.3.32 True or false? An *RD* of 0.7 per 100 indicates a positive association between the exposure and disease.

R.3.33 What happens to the numeric value of the *RR* if we switch designation of the exposed group and nonexposed group?

R.3.34 A risk ratio is 1.85. How much does the exposure *increase* risk in relative terms?

R.3.35 What statistic quantifies the proportion of cases that would be averted had the exposure been absent in exposed cases?

R.3.36 What statistic quantifies the proportion of cases that would be averted if the exposure is eliminated from the population?

References

Alberg, A.J., Ford, J.G., and Samet, J.M. (2007). Epidemiology of Lung Cancer*
ACCP Evidence-Based Clinical Practice Guidelines, 2nd edn, *Chest*, **132**, 29S–55S.

Alho, J.M. (1992). On prevalence, incidence, and duration in general stable populations. *Biometrics*, **48** (2), 587–592.

Blair, S.N., Kohl III, H.W., Barlow, C.E., Paffenbarger, R.S., Jr, Gibbons, L.W., and Macera, C.A. (1995) Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *JAMA*, **273**, 1093–1098.

Brownson, R.C., Remington, P.L., and Davis, J.R. (1993). *Chronic Disease Epidemiology and Control*, American Public Health Association, Washington, DC.

CDC (1992) Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. *MMWR Recomm. Rep.*, **41** (RR-14), 1–7.

Doll, R., and Peto, R. (1976). Mortality in relation to smoking: 20 years' observations on male British doctors. *British Medical Journal*, **2** (6051), 1525–1536.

Doll, R., and Peto, R. (1981). *The Causes of Cancer*, Oxford University Press, New York.

Elandt-Johnson, R.C., and Johnson, N.L. (1980) *Survival Models and Data Analysis*, John Wiley & Sons, New York.

Farrow, D.C. and Thomas, D.B. (1998) Cancer, in *Maxcy-Rosenau-Last Textbook of Public Health and Preventive Medicine*, (ed. R.B. Wallace), Appleton & Lange, Norwalk, CT.

Gladwell, M. (2001, June 11) Wrong Turn, New Yorker, 50–61.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., and Stokes III, J. (1961). Factors of risk in the development of coronary heart disease—six-year follow-up experience. *Annals of Internal Medicine*, **55**, 33–50.

Miller, A.B. (1992). Planning cancer control strategies, in *Chronic Diseases in Canada*, vol. 13, No. 1, *Health and Welfare*, Ontario.

Milunsky, A., Jick, H., Jick, S.S., Bruell, C.L., MacLaughlin, D.S., Rothman, K.J., and Willett, W. (1989). Multivitamin/folic acid supplementation in early pregnancy reduces the prevalence of neural tube defects. *JAMA*, **262**, 2847–2852.

National Center for Health Statistics (NCHS), (1993) *Vital Statistics of the United States, 1991, Volume 11, Mortality, Part B*, U.S. Department of Health and Human Services, Hyattsville, MD, PHS No. 90–1102.

National Center for Health Statistics (NCHS), (1995) *Health, United States, 1994*, U.S. Government Printing Office, Washington, DC, DHHS Publication No. (PHS) 95–1232.

Reynolds, P., Greenberg, R.S., Buffler, P.A., *et al*. (1991) Lung cancer in nonsmoking women: a multicenter case–control study. *Cancer Epidemiology Biomarkers and Prevention*, **1**, 35–43.

Siscovick, D.S., Weiss, N.S., Fletcher, R.H., and Lasky, T. (1984). The incidence of primary cardiac arrest during vigorous exercise, *New England Journal of Medicine*, **311**, 874–877.

Smith, P.F., Mikl, J., Truman, B.I., Lessner, L., Lehman, J.S., Stevens, R.W., Lord, E.A., Broaddus, R.K., and Morse, D.L. (1991) HIV infection among women entering the New York State correctional system. *American Journal of Public Health*, **81** (Suppl.), 35–40.

United Nations Department for Economic and Social Information and Policy Analysis (United Nations) (1996) *1994 Demographic Yearbook*, United Nations, New York, Publication E/F.96.XIII.1.

U.S. Bureau of the Census. (1992). *Statistical Abstracts of the United States 1992* U.S. Government Printing Office, Washington, DC.

U.S. Bureau of the Census (1994). *Statistical Abstracts of the United States 1994* U.S. Government Printing Office, Washington, DC.

U.S. Bureau of the Census (1995). *Statistical Abstracts of the United States (115th edition): 1995*, U.S. Government Printing Office, Washington, DC.

Zhou, Y.F., Leon, M.B., Waclawiw, M.A., Popma, J.J., Yu, Z.X., Finkel, T., and Epstein, S.E. (1996) Association between prior cytomegalovirus infection and the risk of restenosis after coronary atherectomy. *New England Journal of Medicine*, **335**, 624–630.

Addendum: additional mathematical details

1 **Cohorts are closed populations**. A theoretical epidemiologist once said 'once you are a member of a cohort, you are a member for life.' By this he meant that, in theory, cohorts lose members only when study subjects are no longer at risk of becoming a case. Of course, actual cohort studies lose study subjects when participants withdraw from the study or are ''lost to follow-up.'' Fortunately, statisticians have developed methods to compensate for such withdrawals (see Chapter 17). In addition, note that cohort studies may have rolling period of enrollment during which subjects are recruited into the study. Some sources refer to this type of study as an ''open cohort study.'' However, ''open cohort'' is an oxymoron—cohorts are by definition closed populations: once a study subject is ''enclosed'' in a cohort they are followed until they are no longer at risk of becoming a case or the study ends.

2 **Different types of mathematical ratios**. Rates and proportions are types of ratios. A **ratio** is simply a combination of two numbers (a *numerator* and *denominator)* that shows their relative sizes. A ratio can be expressed by separating its numerator and denominator with a colon (*a*:*b*), by representing the relation as a fraction (*a*/*b*), or by completing the division and expressing the result in decimal form (*x.xxx*).
A key to effectively working with a ratio is understanding its numerator and denominator. For example,

$$\text{Body mass index (BMI)} = \frac{\text{weight in kg}}{(\text{height in meters})^2}$$

This index has a numerator representing weight (kg) and a denominator representing ''length squared'' (m^2). The **dimensionality** of a ratio is its combined scale of measure. BMI, for instance, has weight-per-length-squared dimensionality measured in kg/m^2 units. Dimensionality and units of observations may seem arcane, but these concepts can be very helpful in understanding ratios.
The most common types of ratios used in epidemiology are rates, proportions, and odds. A **rate** (λ, lamba) is a specific type of ratio used to quantify dynamic processes such as growth and speed. The general formula for a rate is

$$\lambda = \frac{\Delta A}{\Delta T}$$

where Δ (capital delta) represents ''change in,'' A represents one quantity, and T represents another quantity usually containing an element of time. As a familiar example, speed is a rate at which the numerator is a change in distance (e.g., miles) and the denominator is a change in time (e.g., hours). When combined, ''speed'' has a dimensionality of ''distance per unit time'' (e.g., miles per hour).
A **proportion** is a ratio in which the numerator is a subset of the denominator:

$$p = \frac{A_1}{A_0 + A_1}$$

where A_1 represents the number of elements positive for an attribute and A_0 represents the number of elements negative for this same attribute. For example, if 5 people are positive for a disease and 95 are negative, then the disease proportion

p = 5 people/(5 people + 95 people) = 5 ~~people~~/100 ~~people~~ = 0.05. Notice that the ''people units'' in the numerator and in the denominator of the proportion cancel out upon division leaving a dimensionless number. In addition, since the numerator is a subset of the denominator, proportions have the limited range of 0 to 1.0.

Odds offer an alternative to proportions when working with binary attributes. If A_1 represents the number of ''positives'' and A_0 represents the number of ''negatives,'' then the odds are

$$o = \frac{A_1}{A_0}$$

For example, a sample with 5 people with a disease and 95 without the disease has a disease odds of 5 : 95 = 0.0526.

The relation between an odds and a proportion is simple: $o = p/(1 - p)$ and $p = o/(1 + o)$. For example, when $p = 0.05$, $o = 0.05/(1 - 0.05) = 0.0526$. When $o = 0.0526$, $p = 0.0526/(1 + 0.0526) = 0.05$.

When the proportion being studied is small (say, less than 5%), then the proportion and its analogous odds will be approximately equivalent. However, when the proportion is not small, the odds will exceed its analogous proportion. For example, when $p = 0.5$, $o = 0.5/(1 - 0.5) = 1.0$. Like proportions, odds are dimensionless. However, unlike proportions, odds have an unlimited range of 0 to ∞, and are undefined when their denominator is 0.

3 **Perception of risk not always in line with reality**. The perception of risk is not always in line with actual occurrence. In general, the public is more fearful of small risks they cannot control than large risks they feel as if they can control. Culture and the media have roles in shaping public perceptions of risk, as does the public's misapprehension of numbers. For an introduction this topic, see Slovick, P. (1987) Perception of Risk. *Science*, **236**, 280–285.

4 **Mathematical relationships between incidence rates and incidence proportion**. In Section 3.1 it was stated that when a disease is rare, or at least not common (risk approximately less than 10%), then risk $\approx$ rate $\times$ time. Moreover, when the time of observation is 1 year, this expression simplifies to one-year risk $\approx$ rate per person-year. These two examples will illustrate the point.

(a) Rare outcome. Suppose the risk of death over a year in a cohort of 100 people is 0.01. We therefore expect one death. If we apply an actuarial adjustment in calculating the mortality rate (Section 3.1), then the mortality rate $= \dfrac{1 \text{ death}}{\left(1 \text{ person} \times \frac{1}{2} \text{ year}\right) + (99 \text{ people} \times 1 \text{ year})} = 0.0100503 \text{ year}^{-1}$, and the risk and rate are about equal numerically.

(b) Common outcome. Now suppose that half (50) people die in the cohort of 100. Assuming death occurs uniformly over the interval, the mortality rate $= \dfrac{50 \text{ death}}{\left(50 \text{ people} \times \frac{1}{2} \text{ year}\right) + (50 \text{ people} \times 1 \text{ year})} \cong 0.67 \text{ year}^{-1}$, which is quite different numerically than the risk of 0.5.

(c) When rate is constant. When the rate of disease is constant in a cohort, the population shows an exponential decline and

$$\text{Risk} = 1 - e^{-(\text{rate} \times \text{time})}$$

where e represents the universal constant ($e \approx 2.718\,281$). For instance, a disease that occurs at a constant rate of 0.6667 year^{-1} equates to a 1-year

incidence proportion (risk) of $1 - e^{-(0.6667 \times 1)} = 0.5134$, which is not all that different than what one achieves with the actuarial adjustment (see point 4a, above).

5 **Mathematical relationship between prevalence and incidence**. In Section 3.1 we noted that prevalence $\approx$ (incidence rate) $\times$ (average duration) when the disease is rare. When the disease is not rare, (prevalence odds) $\approx$ (incidence rate) $\times$ (average duration), where the (prevalence odds) = prevalence/(1 − prevalence). However, both of the aforementioned treatments require steady-state assumptions, which are rarely achieved in populations. Therefore, there is no simple formula for the relation between incidence and prevalence that can be used in practice. A complex formula that derives overall prevalence based on age-specific incidences and survival times, coverage of which is beyond the scope of this book, is presented in Alho, J.M. (1992) On prevalence, incidence, and duration in general stable populations. *Biometrics*, **48** (2), 587–592.

6 **Relation between the rate ratio and risk ratio**. We noted in Section 3.2 that the rate ratio will be equivalent to risk ratio in most situations. This is true when the disease is rare. However, when the disease is common, the value of the rate ratio will be more extreme than the value of the risk ratio. Consider as an example a situation in which 50 out of 100 exposed individuals become ill over the course of the year ($\text{risk}_1 = 0.50$) and 25 out of 100 nonexposed individuals become ill ($\text{risk}_0 = 0.25$) for a risk Ratio of 0.50/0.25 = 2.00. If we assume the onsets of illness occur approximately uniformly over time, the associated rates will be rate_1 = 50/[(50 healthy people $\times$ 1 year) + (50 disease onsets $\times$ 1/2 year)] = 50/75 person-years = 0.6667 year^{-1} and rate_0 = 25/[(75 $\times$ 1 year) + (25 $\times$ 1/2 year)] = 0.2857 year^{-1} for a rate ratio of 0.6667/0.2857 = 2.33. In fact, it can be shown that the rate ratio will be equivalent to the odds ratio whether the disease is rare or common.

CHAPTER 4
Descriptive Epidemiology

4.1 Introduction
- What is descriptive epidemiology?
- Case series
- Surveillance systems
- National health surveys and vital record systems

4.2 Epidemiologic variables
- Person
- Place
- Time

4.3 Ecological correlations
- Aggregate-level data
- The ecological fallacy
- Other types of aggregate-level variables

Exercises

Review questions

References

I keep six honest serving men
(They taught me all I know);
Their names are what and why and when
And how and where and who

Rudyard Kipling

4.1 Introduction

What is descriptive epidemiology?

Descriptive epidemiology is a general term used to refer to a broad array of epidemiologic activities whose primary purpose is to describe disease occurrence and generate hypotheses and ideas about cause. Traditionally, this subject has been taught in terms of describing disease occurrence according to the **epidemiologic variables** of **person**, **place**, and **time**.

In contrast to descriptive epidemiology, **analytic epidemiology** starts with specific hypotheses about cause and then designs its studies to address these specific hypotheses. Analytic epidemiologic studies address specific hypotheses, while descriptive

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

epidemiologic studies are more exploratory or ''hypothesis generating.'' With this said, it should be noted that there is no firm demarcation between descriptive epidemiology and analytic epidemiology: all epidemiologic studies serve to advance knowledge of disease causation and prevention, and many studies serve both descriptive and analytic purposes. From a learning perspective, however, it remains useful to note that studies tend to fall toward one end or the other of the descriptive-analytic spectrum.

Descriptive epidemiology often uses data from standing sources (i.e., routinely collected data). Three such sources are case series, surveillance systems, vital records, and nation health surveys.

Case series

Case series describe the medical history and clinical manifestations of a small number of individuals with a particular disease or syndrome. ''Denominator data'' is absent from case series. Therefore, case series cannot calculate incidence or prevalence. In addition, no referent or ''control'' series is present. Therefore causal conclusions are often beyond the reach of case series analysis.[a] Nevertheless, observations derived from cases series often signal an emerging problem and help clarify hypotheses for further investigation. An example follows.

Illustrative Example 4.1 Acquired immune deficiency syndrome (case series)

In 1981, local clinicians and the Epidemic Intelligence Service Officer stationed at the Los Angeles County Department of Public Health prepared and submitted a report of five cases of *Pneumocystis* pneumonia in previously healthy young men (CDC, 1981, 2001). Before publication, editorial staff at the CDC sent the report to experts in parasitic and sexually transmitted diseases who noted that the case histories suggested that they were dealing with cellular-immune dysfunction disease acquired through sexual contact. At about the same time, the sole distributor of the antifungal drug (pentamidine) used to treat *Pneumocystis* pneumonia in the United States began receiving multiple requests for the medicine from physicians throughout the country. The affected individuals were, again, young men. In June 1981, CDC developed an investigative team to develop a case definition and identify risk factors for this new syndrome. Within a couple of years, a case definition for acquired immunodeficiency syndrome (AIDS) had been established and major risks factors for the condition had been identified.

Surveillance systems

Epidemiologic **surveillance systems** are structures set up to routinely collect and analyze data for specific types of health outcomes. Epidemiologic surveillance systems may be either active or passive in nature. **Active surveillance systems** require actively seeking-out cases in defined populations, and thus requires the use of specially trained personnel to retrieve and review health care and laboratory records to discover and confirm cases. In contrast, **passive surveillance** relies on health professionals and the public to identify cases and submit reports to the surveillance system.

[a] In rare instances, when the physiological, clinical, and supporting evidence is compelling, causal conclusions are possible from cases series or even from just a single case

Many different types of surveillance systems exist. Three examples are the Surveillance, Epidemiology and End Results program for monitoring cancer occurrence and treatment, the National Notifiable Diseases Surveillance System for monitoring reportable diseases, and the Food and Drug Administration's MedWatch system for monitoring food and drug safety.

The **Surveillance, Epidemiology and End Results (SEER)** program of the National Cancer Institute is an active surveillance system that functions as the primary source of cancer statistics in the United States. SEER registries routinely collect data on patient demographics, primary tumor site, tumor morphology, stage at diagnosis, first course of treatment, and patient survival. Data from the Census Bureau are used as denominator information to calculate cancer rates within the capturement area of each of the SEER registries. SEER then compiles cancer statistics from each region to estimate cancer incidence for the entire country.

Illustrative Example 4.2 Endometrial cancer (active surveillance)

Figure 4.1 shows sharp rises in uterine cancer incidence in five regions of the United States between 1969 and 1973 (Weiss *et al*., 1979). Data within regions demonstrate increases of more than 10% per year over the period of observation. When the investigators further scrutinized these data, they found that the sharpest increases were among middle-aged women (data not shown). The investigators also noted that these increases paralleled large scale increases in the prescribing of estrogen for symptoms of menopause and osteoporosis that occurred concurrently, leading to a hypothesis that unopposed estrogen may increase the risk of endometrial cancer in middle-aged women. Analytic epidemiologic studies that followed this lead confirmed the association. In addition, studies in laboratory animals showed that estrogen stimulated cell proliferation of the inner lining of the uterus. Thus, the initial hypothesis raised by descriptions of increased rates were corroborated, leading to discontinuing the use of unopposed estrogen (estrogen without progestin) in post-menopausal women with intact uteri.

Aside: Some individuals may have the mistaken impression that the rates in Figure 4.1 are longitudinal. However, rates derived from open populations do not

Figure 4.1 Figure for Illustrative Example 4.2 Endometrial cancer rates in five regions of the USA, 1969–1973. (Based on data from Weiss *et al*., 1976).

follow individual experience over time. Therefore, these data represent a series of ''current'' or ''cross-sectional'' rates—see Chapter 3 and Chapter 18 for additional information about open population current incidence rates.

The **National Notifiable Disease Surveillance System** provides another example of a surveillance system. This system collects reports of selected infectious and noninfectious notifiable diseases in the United States. The list of notifiable diseases is updated regularly by state legislation in collaboration with the U.S. Centers for Disease Control and Prevention.[b] Statistical summaries of reportable diseases are published in each volume of the Morbidity and Mortality Weekly Report (MMWR).

As a third example of a surveillance system, let us consider the U.S. Food and Drugs Administration's **MedWatch** program. Since the Food and Drugs Act of 1906, the U.S. Food and Drug Administration has been the federal agency responsible for protecting the public health through the regulation and supervision of foods, cosmetics, drugs, and medical devices. The FDA instituted the *Medwatch* program in 1993 as a unified system by which consumers and health care professionals can voluntarily report suspected serious adverse events and product quality problems associated with the use of FDA-regulated products. Thus, *Medwatch* is a passive surveillance system. Because *MedWatch* relies on voluntary reports, its listing of cases is often incomplete. In addition, some cases may be ''false positives.'' Thus, like most passive surveillance systems, *MedWatch* is insensitive to subtle changes. Nonetheless, when used judiciously, even passive surveillance system such as the FDA's *MedWatch* system can be useful in signaling problems of emerging public health threats as this example illustrates.

Illustrative Example 4.3 Suprofen-associated flank pain (passive surveillance)

In this example, data from the FDA's forerunner of the *MedWatch* system identified a syndrome of flank-pain and transient renal failure caused by an analgesic medication called suprofen. Figure 4.2 plots the number of flank-pain syndrome cases reported to the FDA by month of onset (open bars) and month of report (solid bars), along with marketing data for the drug (dashed-line). The passive surveillance system was stimulated by ''Dear Doctor'' letters, indicated by the arrows on the graph, which alerted all US physicians of the emerging problem. Identification of this unanticipated adverse reaction ultimately led to withdrawal of the drug from the market by the drug manufacturer in 1987 (Rossi *et al*., 1988).

National health surveys and vital record systems

Governments, as part of their responsibility to monitor the health of their populations, routinely collect data on births, deaths, and various health parameters. **Birth certificates** are used to calculate birth rates and rates of conditions that affect the perinatal period, such as congenital malformations, birth weight, length of gestation, fetal death, and demographic characteristics of the parents. **Death certificates** are completed by funeral directors and attending physicians to include demographic information about the decedent and information about their cause of death. Deaths that are accident-, suicide-, or homicide-related are completed by the medical examiner or coroner as

[b] For a list of the current reportable diseases see the National Notifiable Diseases Surveillance System website *www.cdc.gov/osels/ph_surveillance/nndss/nndsshis.htm*.

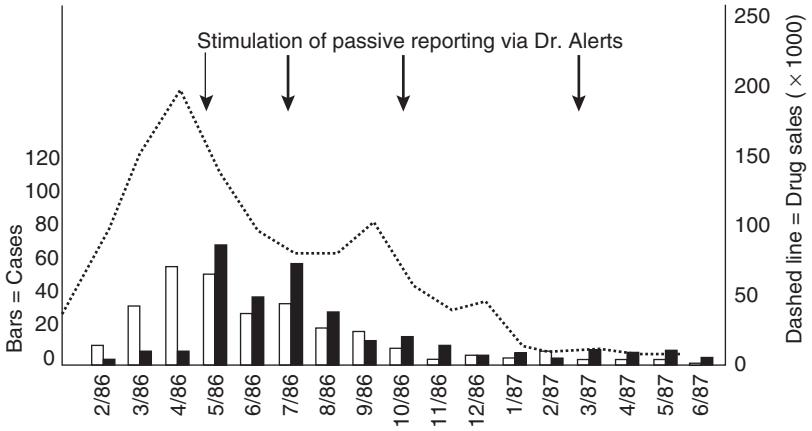


Figure 4.2 Figure for Illustrative Example 4.3 Reports of flank pain and marketing data for the analgesic suprofen (Data from Rossi *et al.*, 1988).

part of the investigation of the cause of death. (Local laws dictate which deaths a coroner must investigate.) In the USA, state and local registrars check the information on birth and death certificates for completeness and accuracy before forwarding copies to the National Center for Health Statistics (NCHS) for recording and compilation. Birth and mortality statistics are compiled and published in various publications, such as *Vital Statistics of the United States* and *Health, United States*. (See *http://www.cdc.gov/nchs/* for a list of publications.)

In addition to tracking births and deaths, nations routinely maintain **health surveys** to track levels of diseases and disease determinants in populations. These surveys also include information about bodily characteristics, behavior, nutrition, health care, and other health concerns of the citizens. In the USA, the agency primarily responsible for compiling these data is the *National Center for Health Statistics* (*www.cdc.gov/nchs*). In Canada, the comparable agency is *Statistics Canada* (*www.statcan.gc.ca*). In Great Britain, the *Office of National Statistics* (*www.statistics.gov.uk*) compiles health statistics. Each of these national agencies maintains multiple health databases. Examples of survey data from the *U.S. National Center for Health Statistics* are the National Health and Nutrition Examination Survey (NHANES), National Health Interview Survey (NHIS), the National Hospital Discharge Survey (NHDS), National Ambulatory Medical Care Survey (NAMCS), and the National Hospital Ambulatory Medical Care Survey (NHAMCS). The methods employed by these data systems evolve over time and are documented on *www.cdc.gov/nchs/*.

4.2 Epidemiologic variables

Once descriptive epidemiologic data are procured, disease occurrence is tallied according to available person, place, and time variables. **Person variables** address characteristics and attributes of population and population subgroups. **Place variables** are characteristics of the locale in which people live, work, and visit. **Time variables** address disease occurrence in relation to various time parameters such

Table 4.1 Examples of person variables.

Age	Alcohol use
Sex	Body mass index
Ethnicity/race	Host responses to social and environmental stressors
Genetic predispositions	Educational level
Physiologic states (e.g., pregnancy)	Socioeconomic status
Concurrent disease	Occupation
Immune status	Customs
Physical activity	Religion
Marital status	Foreign birth
Dietary practices	Knowledge, attitudes and beliefs
Tobacco use	

as time since exposure, calendar time, and seasonality. Let us start by considering ''person variables.''

Person

Variations in disease rates by person variables provide insights into exposures to agents and differences in host susceptibility. Table 4.1 lists examples of person variables. Two of the more common person variables are age and sex (gender), as addressed by this illustration.

Illustrative Example 4.4 Sports-related injuries

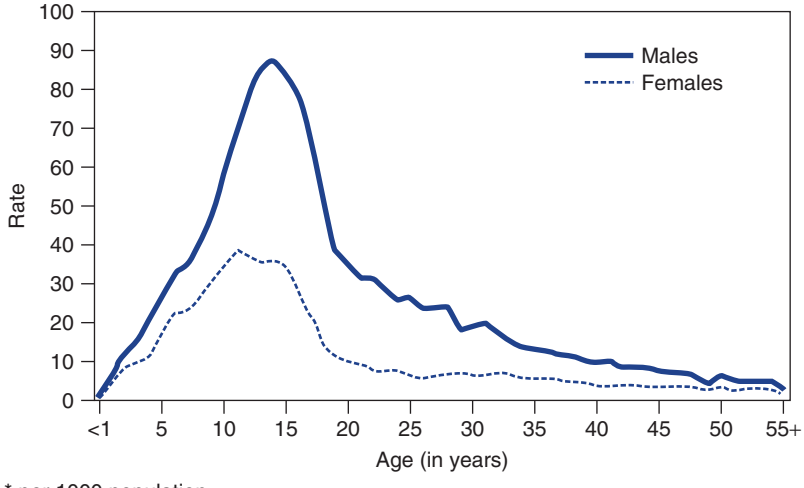
Figure 4.3 displays the age and sex distribution of nonfatal sports- and recreation-related injuries treated in emergency departments for the period July 2000 to June 2001. Rates are highest in males between the ages of 10 and 24, suggesting that special efforts to reduce injuries should be directed toward young males.

Two other common person variables are race and ethnicity. These factors are often related to genetic tendencies, the living habits of individuals, and the level and intensity of various social, biological, and physical environmental exposures.

Illustrative Example 4.5 Tuberculosis among African Americans

Although African–Americans comprise 12% of the US population, they accounted for 33% of the tuberculosis cases reported in 1997. Twenty-three percent (23%) of the tuberculosis cases were in Hispanics and 19% were Asians and Pacific Islanders, even though these groups comprised 11 and 3.5% of the population, respectively (CDC, 2000). High rates of tuberculosis in these three groups can be explained in terms of known risk factors such as birth in a country where tuberculosis is common, HIV infection, and exposure to high-risk settings such as nursing homes, correctional facilities, and homeless shelters.

A person's occupation is an important health determinant. People spend much of their life at work where they are exposed to chemical, physical, biological, and social



* per 1000 population.

Figure 4.3 Rates of nonfatal unintentional sports- and recreation-related injuries treated in emergency departments by age and sex, USA, July 2000–June 2001 (Source: CDC, 2002).

stressors. Occupation is also highly correlated with socioeconomic status and specific constitutional tendencies, all of which have an influence on health.

Illustrative Example 4.6 Brewing beer as a protective factor against cholera

One of the founding members of the London Epidemiological Society, William Augustus Guy (1810–1885), made this insightful observation about the rarity of cholera among brewery workers (Snow, 1855, p. 124):

> . . . the brewers' men seem to have suffered very lightly both in that and the more recent [cholera] epidemics. The reason of this probably is, that they never drink water, and are therefore exempted from imbibing the cholera poison in that vehicle.

Work in the brewing industry, in this instance, proved to be salubrious.

Table 4.2 Host and environmental factors associated with place.

Presence and level of agents	Nutritional practices
Presence of vectors that facilitate transmission	Occupations
Socioeconomic differences	Recreational practices
Genetic characteristics of residents	Urban/rural differences
Physiologic and anatomic attributes of residents	Economic development
Geology	Social disruptions (e.g., war, natural disasters, economic downturns)
Climate	Social norms in behavior
Population density	Medical practices
	Access to health care

Place

Place variables are characteristics of the locale in which people live, work, and visit. Place variables may be defined in terms of geographic boundaries (e.g., street, city, state, region, country) or environmental characteristics (e.g., rural/urban, domestic/foreign, institutional/noninstitutional). Table 4.2 lists examples of host and environmental characteristics associated with place.

Differences in the incidence and prevalence of disease by place are related to differences in host susceptibility or prevalence of causal agents.

Illustrative Example 4.7 Breast cancer mortality

Figure 4.4 compares international breast cancer mortality rates for the period 1958–1959. At that time these data were collected, Japan's breast cancer mortality rate was one-quarter to one-half that of the other countries listed. This raised questions about genetic and environmental contributors to breast cancer. Many hypotheses were generated to address whether the differences were attributable to genetics or environment differences.

Since that time, studies in the USA have demonstrated that breast cancer rates in Japanese–American women increase over successive generations, suggesting a strong environmental component to breast cancer (Buell, 1973). In addition, breast malignancies in Japan have risen over time as the Japanese diet and lifestyle has been progressively westernized (Wynder *et al*., 1991). Environmental theories that have been put forward to explain low rates of breast cancer in mid-20th century Japanese women include the lengthy breast-feeding and long lactation periods among traditional Japanese women (Lilienfeld, 1963), the low body weights of Japanese women (De Waard *et al*., 1977), dietary differences (Armstrong and Doll, 1975), age at menarche (Henderson and Bernstein, 1991), and menstrual cycle length (Wang *et al*., 1992). It is interesting to note that hypotheses regarding diet (e.g., high fat diets) have generally *not* been corroborated.

Mapping can be helpful when exploring patterns of disease occurrence. John Snow's celebrated map of the cholera deaths surrounding the Broad Street pump provides an historical illustration (Figure 1.13). Less celebrated but of no less importance during John Snow's investigations of cholera were his maps of water distribution in Victorian London.

Illustrative Example 4.8 Victorian water pipes

During the 19th century, drinking water in London was supplied by private companies via networks of pipes. The two main suppliers of water in the epidemic areas of London in Victorian times were the Southwark & Vauxhall Water (S&V) Company and the Lambeth Water Company. Figure 4.5 is a section of one of Snow's maps showing water distribution networks in 1849 London. The map has hatched and cross-hatched areas supplied by S&V Company, the Lambeth Company, and areas in which the pipes of both companies were intermingled. By using publically available mortality data, Snow demonstrated that rates of cholera were highest in the areas supplied by the S&V company, lowest in the areas served by the Lambeth Company, and intermediate in the areas of mixed usage. This supported the theory that the S&V Company was disseminating unsafe cholera-laden water.

Time

The occurrence of disease over time can be analyzed from multiple time-perspectives, some of which are presented in Table 4.3.

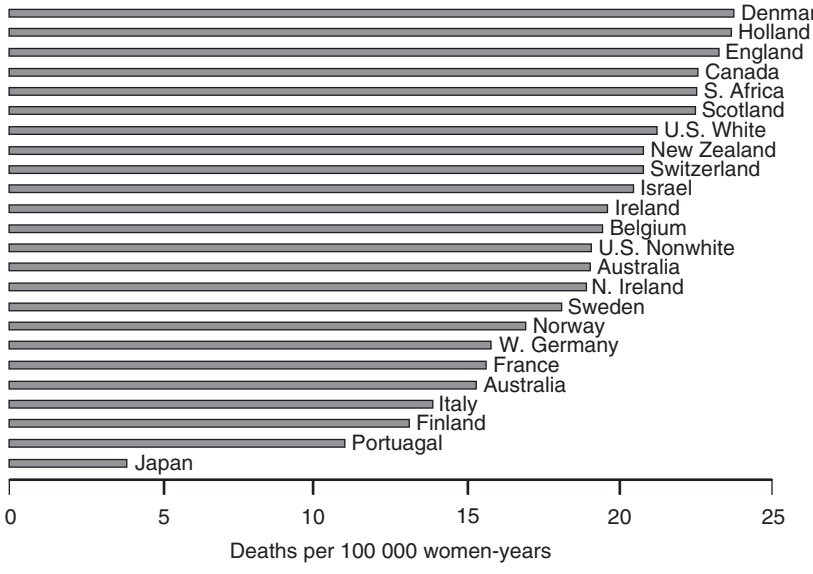


Figure 4.4 Figure for Illustrative Example 4.7. Age-adjusted breast cancer mortality per 100 000 women, 23 countries, 1958–1959 (Based on data in Segi and Kurihara, 1962, p. 31).

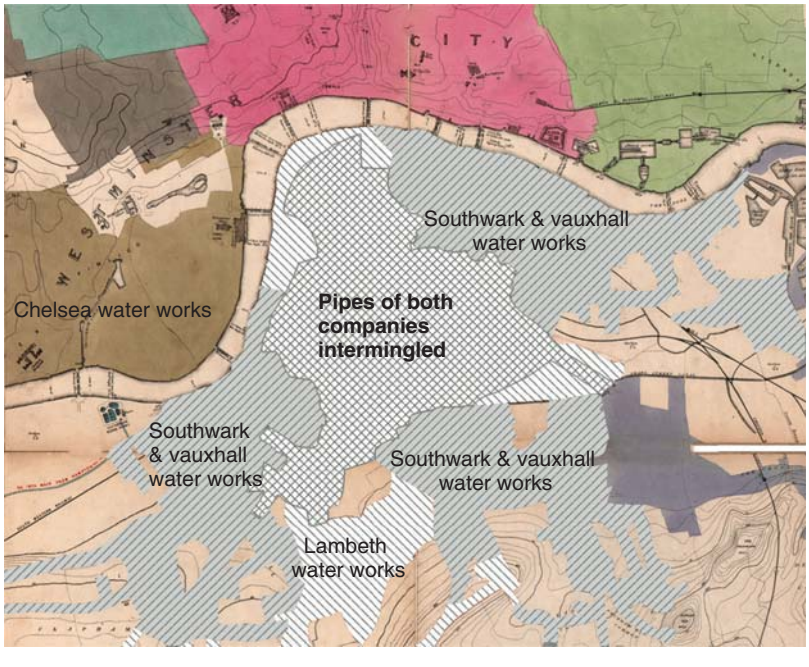


Figure 4.5 A section of John Snow's map showing the distribution of water pipes in 19th century London (Markup of a map in the 1936 reprint of Snow, 1855).

Table 4.3 Examples of time factors.

Calendar time	Time under observation
Time since birth (age)	Time since diagnosis
Time since first exposure	Circadian and other physiological rhythms
Total exposure time	Seasonality

A common way to explore the distribution of cases over time is in the form of an **epidemic curve**. The *Y* axis of an epidemic curve represents the number or percent of cases that occurred during the time-interval indicated on the *X* axis. Epidemic curves provide insight into the induction period of disease and the temporal course of disease occurrence. Figure 4.6 demonstrates these temporal patterns:

(**A**) **Sporadic** (occurring rarely and without regularity)
(**B**) **Endemic** (occurring predictably with only minor or predictable variation)
(**C**) **Point epidemic** (occurring in clear excess over a period of time and then rapidly returning to normal)
(**D**) **Propagating epidemic** (occurring in clear excess with continuing increases over time).

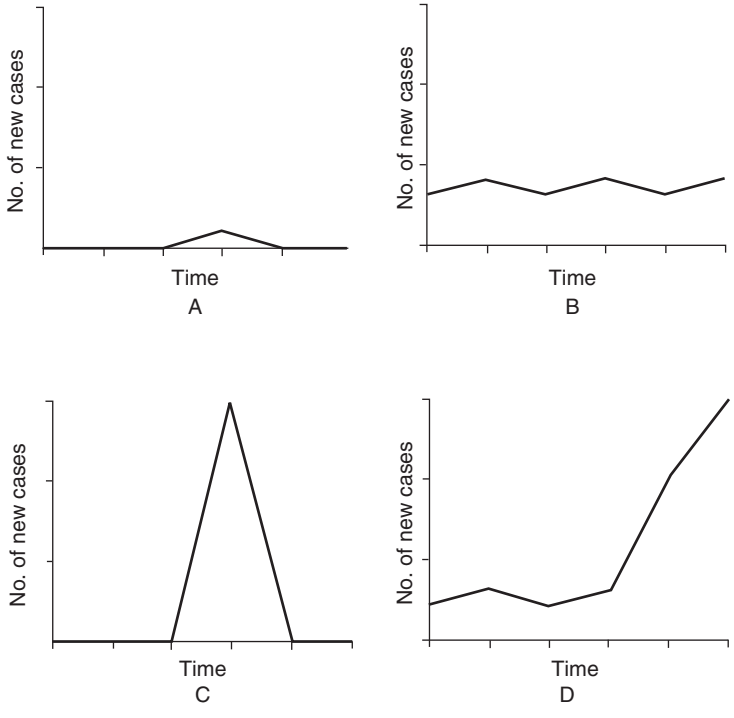


Figure 4.6 General patterns of occurrence: (A) sporadic, (B) endemic, (C) point epidemic, and (D) propagating epidemic.

Illustrative Example 4.9 Golden Square epidemic curve

Figure 4.7 shows the epidemic curve of the historically important Golden Square 1854 cholera epidemic investigated by John Snow (see Section 1.3). John Snow did not produce an epidemic curve during his investigation, but Bradford Hill's did, some 100 years later (Hill, 1955) (Figure 4.7). Scrutiny of this curve reveals that the epidemic was on the decline when the handle of the infamous Broad Street pump was removed on September 8. This suggests that removal of the pump handle was *not* decisive in ending the epidemic—the epidemic was apparently in the process of ''burning itself out of susceptibles'' by the time the pump handle was removed.

Figure 4.8 exhibits the seasonal fluctuations for pneumonia and influenza in the USA between 2006 and 2010. The seasonal baseline and epidemic threshold are displayed as parallel wavy-lines. When the observed number of cases exceeds the epidemic threshold for two consecutive weeks, further investigation is undertaken. Notice that epidemic thresholds were broken during the first part of 2008 and toward the end of 2009.

Illustrative Example 4.10 Tuberculosis trends

Figure 4.9 plots tuberculosis rates from 1953 to 2008 in the USA. In 1953, when nationwide tuberculosis reporting first began, there were more than 84,000 tuberculosis cases reported annually for a rate of 52.6 per 100 000 person-years. From 1953 through 1985, the rate of tuberculosis dropped precipitously. Between 1985 and 1992, however, there was a modest increase. This increase was traced to the HIV epidemic, increases in immigration from countries where tuberculosis was endemic, and increases in the transmission of tuberculosis in high-risk environments such as homeless shelters. In 1993, the upward trend was reversed and the downward trend resumed.

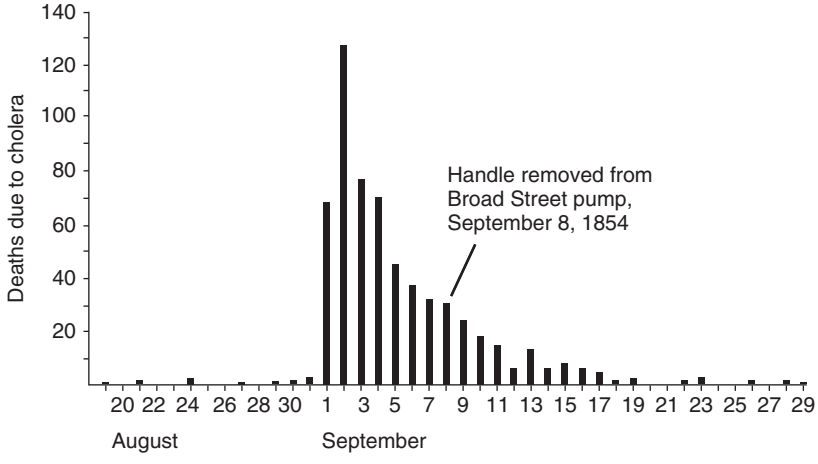


Figure 4.7 Figure for Illustrative Example 4.9. Epidemic curve of the London Golden Square cholera epidemic of 1854 (Data from Snow, 1855, p. 49).

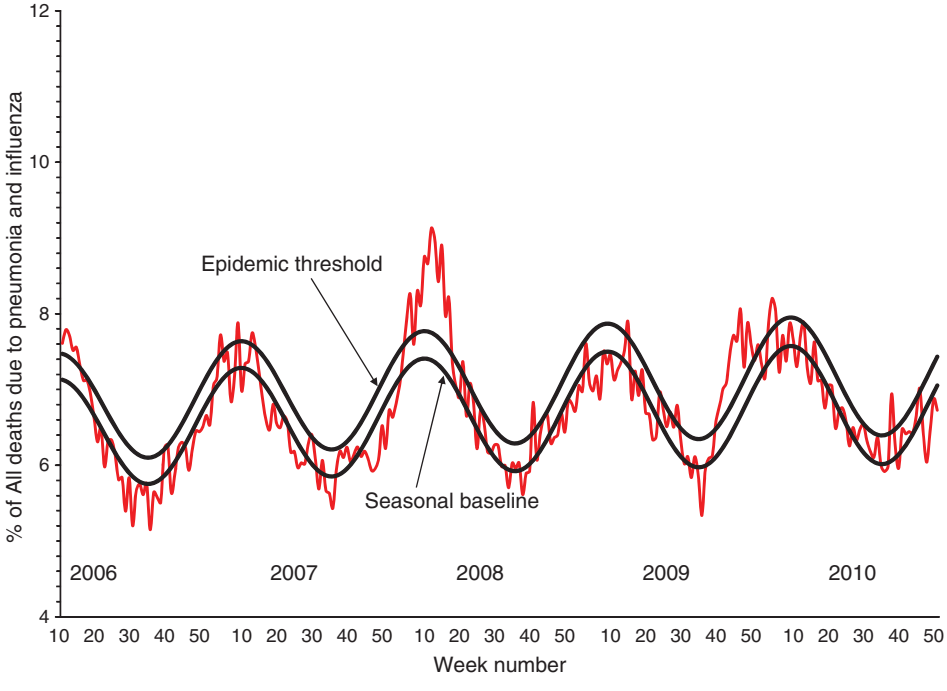


Figure 4.8 Pneumonia and influenza mortality for 122 US cities for 2006–2010 *(Source: http://www.cdc.gov/flu/weekly/pdf/External_F1050.pdf)*.

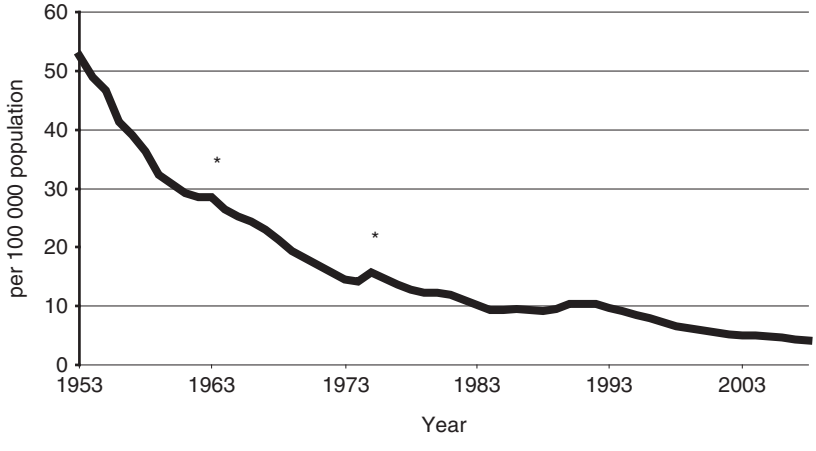


Figure 4.9 Figure for Illustrative Example 4.10. Tuberculosis rates per 100 000 population, USA, 1953–2008. The symbol * indicates change in reporting criteria (Source: CDC, 2009).

4.3 Ecological correlations

Aggregate-level data

The **unit of observation** in a study is the level of aggregation upon which measurements are recorded. In modern epidemiologic studies this is most often at the level of the individual. However, measurements can also be made at the aggregate-level. Epidemiologic studies based on aggregate-level variables are called **ecological studies**.

Ecological studies frequently rely on standing data sources and are often pursued early in our understanding of an epidemiologic problem. Thus, ecological data are often incomplete and lacking information on the multiple factors that contribute to disease occurrence. This is one of the reasons why ecological studies fall toward the descriptive end of the descriptive-analytic spectrum of epidemiologic study designs. Several illustrations are provided.

Illustrative Example 4.11 Smoking and lung cancer (ecological correlation)

Table 4.4 lists aggregate-level data for cigarette consumption and lung cancer mortality in 11 industrialized European nations. The independent variable is per capita cigarette consumption in each region in 1930. The dependent variable is the lung cancer mortality rate per 100 000 in the region in 1950. Figure 4.10 displays these data as a scatterplot, demonstrating a strong positive correlation ($r = 0.74$). Figure 4.11 exhibits parallel ecological increases between cigarette consumption and lung cancer mortality in the UK between 1900 and 1947. These types of ecological analyses provided early support for the smoking–lung cancer hypothesis.

Exploration of ecological correlations also contributed to our early understanding of the relation between dietary fat and coronary artery disease. As early as 1932, Raab noted ''the relative rarity of atherosclerosis and hypertension among the chiefly vegetable-consuming inhabitants of China, Africa and Dutch East Indies and British India ... and the enormous frequency of arteriosclerosis and hypertension among the

Table 4.4 Data for Illustrative Example 4.11.

Country	Per capita cigarette consumption, 1930	Lung cancer mortality per 100 000
United States	1300	20
Great Britain	1100	46
Finland	1100	35
Switzerland	510	25
Canada	500	15
Holland	490	24
Australia	480	18
Denmark	380	17
Sweden	300	11
Norway	250	9
Iceland	230	6

Data source: Doll (1955).

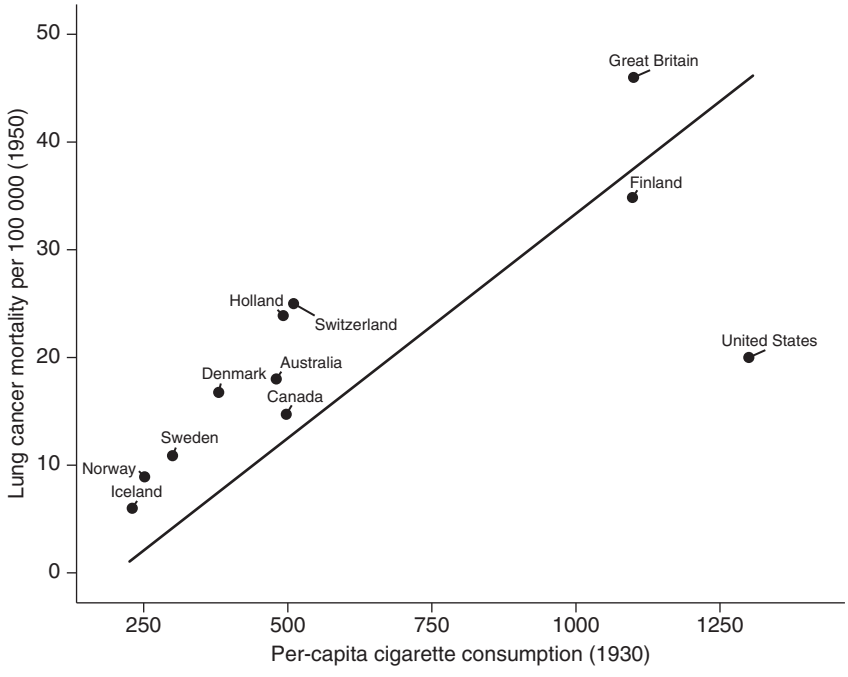


Figure 4.10 Lung cancer mortality (males) in 1950 and per capita consumption of cigarettes in 1930, various countries (Source: Doll, 1955).

people of Europe and North America who consume large quantities of eggs, butter, etc.'' (translated by Stamler, 1989, p. S3).

World War II brought with it prominent reductions in dietary fat consumption, especially in the lands conquered by Nazi Germany. In Norway, where public health and vital statistics were carefully maintained even during the war, there was a clear and prominent decline in mortality from circulatory disease. However, after the war, there was a swift rise in cardiovascular mortality, returning to prewar levels (Strøm and Jensen, 1951).

Illustrative Example 4.12 Dietary fat and cardiovascular disease (ecological correlation)

Figure 4.12 is a replica of a graph from an ecological study by Keys (1953) showing a strong correlation between percentage of calories derived from fat and cardiovascular disease mortality rates in six countries. One criticism that was raised at the time of this study stated that countries with low coronary artery disease mortality and low fat intake differed from high coronary disease countries in ways besides dietary habits, notably in their higher rates of physical activity, lower levels of obesity, and lower rates of smoking. Thus, more refined epidemiologic studies and laboratory studies were needed to sort out the effects of diet, physical activity, genetics, and other elements of the multifactored etiology of coronary artery disease. Dietary fat is now accepted as a valid component cause of coronary disease. However, even today, our understanding about this relation is far from complete. For example, the dose–response relation between specific fatty acids, cholesterol, and coronary heart disease risk have yet to be fully elucidated (Willett, 1990). Thus, the advancement of knowledge about these relationships continues to progress from the initial observation and hypothesis.

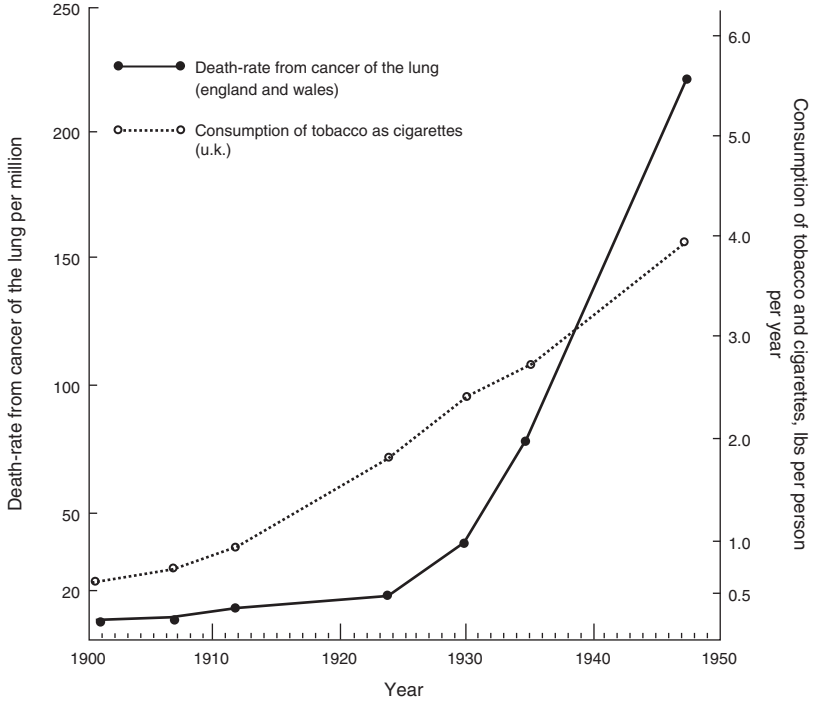


Figure 4.11 Mortality from lung cancer, and cigarette consumption, UK, 1900–1947. The rates are based on 3-year averages for all years except 1947 (Source: Doll and Hill, 1950).

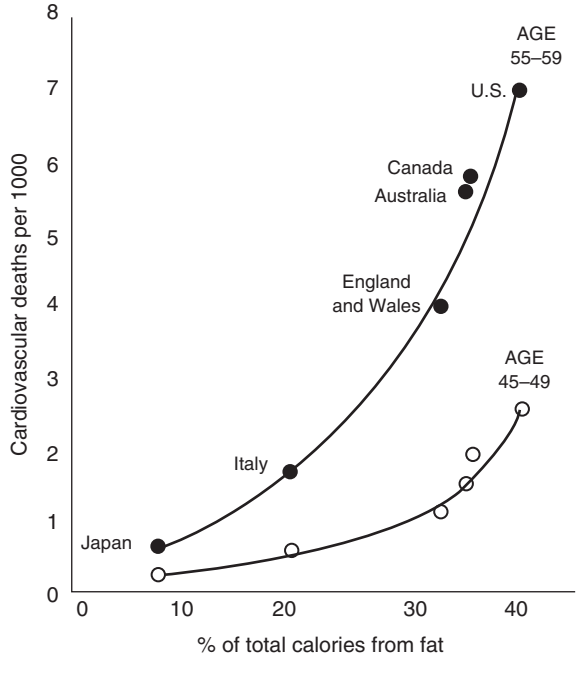


Figure 4.12 Figure for Illustrative Example 4.12. Ecological data: cardiovascular disease mortality and fat calories as a percent of total calories, six countries (Based on Keys, 1953).

The ecological fallacy

The **ecological fallacy** (**aggregation bias**) consists in thinking that an association seen in the aggregate holds true for individuals when in fact it does not (Thorndike, 1939; Selvin, 1958). This is related to the fact that the joint distribution of multiple factors within individuals cannot be teased out in ecological data and that spurious associations remain uncontrolled. The ecological fallacy can thus be viewed as a form of confounding. An historical example follows.

Illustrative Example 4.13 Farr's *faux pas* (ecological fallacy)

William Farr's[c] 1852 study on cholera and geographic altitude in 19th century London provides an opportunity to illustrate an ecological fallacy. At the time, Farr accorded only a small role for contagion as a cause of cholera, placing much greater emphasis on social and environmental conditions (Eyler, 1980). In 1852 Farr wrote: ''Notwithstanding the disturbance produced by the operation of other causes, the mortality from cholera in London bore a certain constant relation to the elevation of the soil, as is evident when the districts are arranged by groups in the order of their altitude.'' Figure 4.13 is a replica of a table from Farr's 1852 paper. This table contains data on altitude above sea level and corresponding cholera mortality rates by neighborhood, along with several other variables (e.g., overall mortality, persons per house).

Figure 4.14 is a scatterplot of these data. The curved line in the Figure 4.14 is the mathematical model Farr proposed to explain the relationship between elevation and cholera mortality. Although the line shows a remarkably good fit to the data, there is no causal relationship between elevation and cholera mortality. The link between the two is that Farr failed to account for the fact that people living at low elevations were more likely to draw drinking water from sources contaminated with *Vibrio cholerae*. Thus, the ecological relationship between elevation and cholera mortality was confounded by proximity to contaminated water sources.

Number of districts	Elevation in feet above trinity high-water mark	Observed average						
		Annual mortality to 10 000 persons living		Number of persons to		Average annual value of		Poor Rate in the £ of house-rent, 1842-43.
		Cholera (1849)	All causes (1838-44)	An acre	A house	Houses.	House and shop room to each person	
						£	£	£
16 ••••	Under 20 ft.	102	251	74	6•8	31	4•645	•072
7 ••••	20–40 ,,	65	237	105	7•6	56	7•358	•071
8 ••••	40–60 ,,	34	235	184	8•5	64	7•342	•056
3 ••••	60–80 ,,	27	236	152	8•8	52	6•374	•049
2 ••••	80–100 ,,	22	211	44	7•7	38	5•183	•036
1 ••••	100 ,,	17	227	102	9•8	71	7•586	•043
1 ••••	350 ,,	8	202	5	7•2	40	5•804	••••
All London ••••••••			252	29		40	5•419	•063

Figure 4.13 Data for Illustrative Example 4.13. London water districts arranged according to their elevation above sea level. (Table that appeared in Farr's 1852 article.)

[c] William Farr (1807–1883), the first registrar of vital statistics in London, is considered one of the founders of epidemiology as practiced in its modern form. See Section 1.3 for a brief biography.

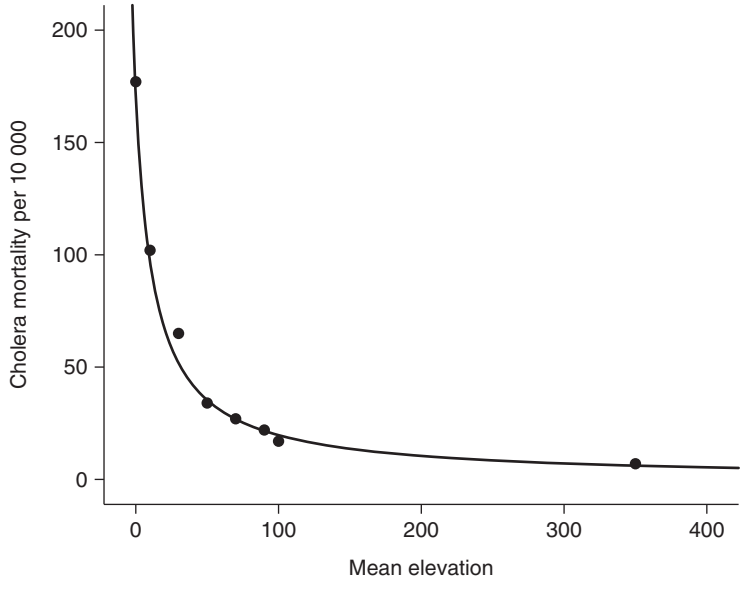


Figure 4.14 Farr's data on cholera mortality and elevation above sea level. The line represents Farr's predictive model: cholera mortality = 2226/(elevation + 13).

Another example of the ecological fallacy is exhibited when one considers rates of coronary heart disease in wealthy and impoverished countries. Wealthy countries demonstrate higher rates of coronary heart disease than poor countries. However, within countries, the relationship is reversed: poor individuals have a higher risk of CHD than wealthy individuals. This paradox is explained by the fact that atherogenic factors such as sedentary lifestyles, obesity, and diets high in fats and calories are more prevalent in wealthy nations.

Other types of aggregate-level variables

Aggregate-level variables we have considered so far are compilations or averages of individual-level characteristics. For example, ''per-capita cigarette consumption'' (Illustrative Example 4.11) represents the average cigarette consumption per person. Similarly, the ecological measurement of ''total calories from fat as a percent of total calories'' (Illustrative Example 4.12) is an average of individual values. However, some aggregate-level variables have no analogous measurement at the individual level. These variables are called **integral group properties** (Susser, 1994). Examples of integral group properties include population density, social disorganization, and existence of specific laws and forms of governance. Exposure to integral group properties are homogeneous within a population, precluding individual-level variation.

In addition **contextual variables** are aggregate-level variables derived from a compilation of individual attributes while having an effect that is beyond the sum of their parts. For example, the percent of a population that is immune to an infectious agent is a contextual variable because when the prevalence of immunity exceeds a certain level, herd immunity decreases the risk of infection beyond that capable of individual immunity.

Contagion variables are another important type of aggregate-level variable. Contagion variables are simultaneously independent variables and dependent variables. Whereas contextual variables are independent variables in their own right, contagion variables are both outcomes (dependent) variables and explanatory (independent) variables that have an influence on future outcomes. For example, the prevalence of HIV in a population is a dependent variable that modifies the probability an individual will come in contact with HIV in the future. Contagion variables apply to infectious events, but may also apply to social and psychological variables that produce their effect as the product of interacting ''contagious'' forces.

The prime justification for measuring integral group properties, contextual variables, and contagion variables is to study health effects within an environmental context that alters outcomes in ways not explicable by studies that focus solely on individual-level variables. Thus, there is contemporary interest in combing individual-level and group-level variables in epidemiology through a process called **multilevel analysis**. Multilevel analysis combines aggregate-level and individual-level variables in a way that seeks to untangling relationships among factors on various levels.

Exercises

4.1 Ecological correlations. Table 4.5 displays a correlation matrix from an ecological study on cigarette consumption and selected cancers by Fraumeni (1968). Based on these data, which cancers are correlated with smoking? Which cancers are correlated with each other? Use these findings to generate relevant causal hypotheses.

Table 4.5 Data for Exercise 4.1. Data for 43 states and the District of Columbia, 1960.

		Cigarettes sold per capita	Bladder cancer deaths per 100 000	Lung cancer deaths per 100 000	Kidney cancer deaths per 100 000	Leukemia deaths per 100 000
Cigarettes sold per capita	Correlation	1	0.704	0.697	0.487	−0.068
	Sig. (2-tailed)		0.000	0.000	0.000	0.659
Bladder cancer deaths per 100 000	Correlation	0.704	1	0.659	0.359*	0.162
	Sig. (2-tailed)	0.000		0.000	0.017	0.293
Lung cancer deaths per 100 000	Correlation	0.697	0.659	1	0.283	−0.152
	Sig. (2-tailed)	0.000	0.000		0.063	0.326
Kidney cancer deaths per 100 000	Correlation	0.487	0.359	0.283	1	0.189
	Sig. (2-tailed)	0.001	0.017	0.063		0.220
Leukemia deaths per 100 000	Correlation	−0.068	0.162	−0.152	0.189	1
	Sig. (2-tailed)	0.659	0.293	0.326	0.220	

Data source: Fraumeni (1968)

4.2 Notifiable conditions. List the current nationally notifiable disease and conditions based on information on the National Notifiable Diseases Surveillance System website *www.cdc.gov/osels/ph_surveillance/nndss/nndsshis.htm*.

4.3 Le Suicide. The French sociologists Emile Durkheim (1858–1917) was known for his compelling scientific approach for studying social phenomena. In his *Rules of Sociological Method* (1895), Durkheim sets forth that (a) social explanations require comparisons, (b) comparisons require classification, and (c) classification requires the definition of those facts to be classified, compared, and ultimately explained. Consistent with these rules, Durkheim warned against *notiones vulgares*—the idea that crudely formed concepts of social phenomena without scientific reflection produce only false knowledge: just as alchemy had preceded chemistry and astrology had preceded astronomy, social reflection merely foreshadows true social science. Durkheim's seminal work *Le Suicide* (1897) considered many potential risk factors for suicide, including psychopathological states, race, heredity, climate, season, imitative behavior, religion, social instability, and a host of other social phenomena. Table 4.6 is based on Table XXI from *Le Suicide*.

(A) Based on these data, list four observations about the potential effects of marriage on suicide.

(B) Present an alternative explanation, other than the effect of marriage, for the associations you noted in part (A).

Table 4.6 Rates per million per year, and relative risks of suicide according to marital status, France, 1889–1891.

Ages	Rates per million per year			Relative risks	
	Unmarried	Married	Widowed	Unmarried with ref. to married	Unmarried with ref. to widowed
Men					
15–20	113	500	—	0.22	—
20–25	237	97	142	2.40	1.66
25–30	394	122	412	3.20	0.95
30–40	627	226	560	2.77	1.12
40–50	975	340	721	2.86	1.35
50–60	1434	520	979	2.75	1.46
60–70	1768	635	1166	2.78	1.51
70–80	1983	704	1288	2.81	1.54
Above 80	1571	770	1154	2.04	1.36
Women					
15–20	79.4	33	333	2.39	0.23
20–25	106	53	66	2.00	1.60
25–30	151	68	178	2.22	0.84
30–40	126	82	205	1.53	0.61
40–50	171	106	168	1.61	1.01
50–60	204	151	199	1.35	1.02
60–70	189	158	257	1.19	0.77
70–80	206	209	248	0.98	0.83
Above 80	176	110	240	1.60	0.79

Review questions

R.4.1 Discuss how descriptive epidemiology differs from analytic epidemiology.

R.4.2 True or false? There is a firm demarcation between descriptive epidemiology and analytic epidemiology. Explain your response.

R.4.3 What is a case series?

R.4.4 Why are we unable to calculate rates based on case series?

R.4.5 True or false? Emerging health threats are often first identified by the unplanned observations of an astute clinician.

R.4.6 What is an epidemiologic surveillance system?

R.4.7 Distinguish between active surveillance and passive surveillance.

R.4.8 What federal agency in the United States is primarily responsible for compiling statistics on the nation's health?

R.4.9 Classically, descriptive epidemiology addresses the distribution of disease according to person, _____, and _____ variables.

R.4.10 Is *occupation* a person, place, or time variable?

R.4.11 Provide examples of host factors that are closely tied to ''place.''

R.4.12 Provide examples of environmental factors that are closely tied to ''place.''

R.4.13 How did the breast cancer studies in Japanese–American women described in the text prove that breast cancer has strong environmental causal components?

R.4.14 Match each term with its description.
Terms: endemic, sporadic, point epidemic, propagating epidemic.
Descriptions:
(a) Occurring in clear excess of normalcy with continuing increases over time.
(b) Occurring in clear excess of normalcy; then rapidly returning to normal levels.
(c) Occurring predictably with minor or predictable fluctuations.
(d) Occurring rarely, without regularity.

R.4.15 What is a ''unit of observation''?

R.4.16 Fill in the blank: Epidemiologic studies based on aggregate-level units of observation are called _____ studies.

R.4.17 True or false? Once data are recorded on an aggregate-level, they cannot be disaggregated to reveal an individual's condition.

R.4.18 What is an ecological correlation?

R.4.19 Is ''neighborhood crime rate'' a person-level variable or aggregate-level variable?

R.4.20 True or false? Additional studies are often needed to sort out the effects of factors identified through ecological correlations.

R.4.21 What is the ecological fallacy?

R.4.22 What is a multilevel analysis?

R.4.23 What is confounding?

R.4.24 Match each term with its description.
Terms: integral group property, contextual variable, contagion variable
Descriptions:
(a) A variable derived from a compilation of individual attributes that has an effect that is beyond the sum of its individual parts.
(b) An aggregate-level variable that affects virtually all members of a group.
(c) An aggregate-level variable that is an outcome affecting future occurrences.

R.4.25 What type of variable is ''whether a nation has a written constitution?''

References

Armstrong, B., and Doll, R. (1975) Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *International Journal of Cancer*, **15**, 617–631.

Buell, P. (1973) Changing incidence of breast cancer in Japanese-American women. *Journal of the National Cancer Institute*, , 1479–1483.

CDC (1981) Pneumocystis pneumonia—Los Angeles. *MMWR*, **30** (21), 250–252.

CDC (2000) Epidemiology of Tuberculosis. *http://www.cdc.gov/nchstp/tb/pubs/ssmodules/* (accessed January 2001).

CDC (2001) First report of AIDS. *MMWR*, **50** (21), 429.

CDC (2002) Nonfatal sports- and recreation-related injuries treated in emergency departments–United States, July 2000–June 2001. *MMWR*, **51**, 736–740.

CDC (2009) *Reported Tuberculosis in the United States, 2008*. Atlanta, GA: U.S. Department of Health and Human Services. Available: *http://www.cdc.gov/tb/statistics/reports/2008/default.htm*

De Waard, F., Cornelis, J.P., Aoki, K., and Yoshida, M. (1977) Breast cancer incidence according to weight and height in two cities of the Netherlands and in Aichi prefecture, Japan. *Cancer*, **40**, 1269–1275.

Doll, R., and Hill, A.B. (1950) Smoking and carcinoma of the lung. *British Medical Journal*, **2**, 739–748.

Doll, R. (1955) Etiology of lung cancer. *Advances in Cancer Research*, **3**, 1–50.

Eyler, J.M. (1980) The conceptual origins of William Farr's epidemiology: Numerical methods and social thought in the 1830s, in *Time, places, and persons*, (ed. A.M. Lilienfeld), Johns Hopkins University Press, Baltimore, pp. 1–21.

Farr, W. (1852) Influence of elevation on the fatality of cholera. *Journal of Statistical Society of London*, **15**, 155–183.

Fraumeni, J.F., Jr., (1968) Cigarette smoking and cancers of the urinary tract: geographic variation in the United States. *Journal of the National Cancer Institute*, **41** (5), 1205–1211.

Henderson, B.E., and Bernstein, L. (1991) The international variation in breast cancer rates: An epidemiological assessment. *Breast Cancer Research and Treatment*, **18** (Suppl. 1), S11–S17.

Hill, A. B. (1955) Snow; an appreciation. *Proceedings of the Royal Society of Medicine*, **48**(12), 1008–1012.

Keys, A. (1953) Atherosclerosis: A problem in newer public health. *Journal of Mt Sinai Hospital*, **20**, 118–139.

Kipling, R. (1912) The Elephant's Child, in *Just So Stories*, Doubleday Page & Company, New York.

Lilienfeld, A.M. (1963) The epidemiology of breast cancer. *Cancer Research*, **23**, 1503–1513.

Raab, W. (1932) Alimentare Foktoren in der enstenhung von arteriosklerose and hypertonie. *Medizinische Klinik Supplement*, **28**, 487–521.

Rossi, A.C., Bosco, L., Faich, G.A., Tanner, A., and Temple, R. (1988) The importance of adverse reaction reporting by physicians. Suprofen and the flank pain syndrome. *JAMA*, **259** (8), 1203–1204.

Segi, M., and Kurihara, M. (1963) *Trends in Cancer Mortality for Selected Sites in 24 Countries*, 1950–1959, Department of Public Health, Tohoku University School of Medicine, Sendai, Japan.

Selvin, H. (1958) Durkheim's suicide and problems of empirical research. *American Journal of Sociology*, **63**, 607–619.

Snow, J. (1849) On the pathology and mode of communication of cholera. *London Medical Gazette*, **44**, 745–752.

Snow, J. (1855) On the mode of communication of cholera (1936 Reprint), *Snow on cholera, being a reprint of two papers by John Snow, M.D., together with a biographical memoir by B.W. Richardson, M.D., and an introduction by Wade Hampton Frost, M.D.*, The Commonwealth Fund, New York.

Stamler, J. (1989) Opportunities and pitfalls in international comparisons related to patterns, trends and determinants of CHD mortality. *International Journal of Epidemiology*, **18**, S3–S18.

Strøm, A., and Jensen, A.R. (1951) Mortality from circulatory diseases in Norway 1940–45. *Lancet*, **260**, 126.

Susser, M. (1994) The logic in ecological: I. The logic of analysis. *American Journal of Public Health*, **84**, 825–829.

Thorndike, E. L. (1939) On the fallacy of imputing and correlations found for groups to the individuals or smaller groups composing them. *American Journal of Psychology*, **52**, 122–124.

Wang, Q.S., Ross, R.K., Yu, M.C., Ning, J.P., Henderson, B.E., and Kimm, H.T. (1992) A case–control study of breast cancer in Tianjin, China. *Cancer Epidemiology, Biomarkers and Prevention*, **1**, 435–439.

Weiss, N.S., Szekely, D.R., and Austin, D.F. (1976) Increasing incidence of endometrial cancer in the United States. *The New England Journal of Medicine*, **294** (23), 1259–1262.

Weiss, N.S., Szekely, D.R., English, D.R., and Schweid, A.I. (1979) Endometrial cancer in relation to patterns of menopausal estrogen use. *JAMA*, **242** (3), 261–264.

Willett, W. (1990) *Nutritional Epidemiology*, Oxford University Press, New York.

Wynder, E.L., Fujita, Y., Harris, R.E., Hirayama, T., and Hiyama, T. (1991) Comparative epidemiology of cancer between the United States and Japan. A second look. *Cancer*, **67**, 746–763.

CHAPTER 5
Introduction to Epidemiologic Study Design

5.1 Etiologic research
- Hypothesis statement
- Variables
- Data

5.2 Ethical conduct of studies involving human subjects

5.3 Selected study design elements
- Necessity of a referent (''control'') group
- Experimental versus observational study designs
- Unit of observation
- Longitudinal versus cross-sectional observations
- Cohort versus case–control samples

5.4 Common types of epidemiologic studies

Exercises

Review questions

References

5.1 Etiologic research

> Hypotheses are nets: only he who casts will catch.
>
> *Novalis*

Hypothesis statement

The goal of analytic epidemiology is to clarify causal relations between various determinants (''exposures'') and health outcomes. Advancement in knowledge is an ongoing process that progresses, often gradually, in overlapping stages. An epidemiologic issue may be brought to bear by an intriguing case report. This may be followed by descriptive epidemiologic studies. As hypotheses are generated and refined, more detailed studies follow, perhaps culminating in an epidemiologic experiment. In all instances, the epidemiologist pursues methods that are most advantageous for revealing causal relations.

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

Not to be overlooked in this process is the need to have sharply focused research questions and hypotheses. The study must then be designed in a manner that will connect the data to these questions and hypotheses. Hypotheses must therefore clearly define:

1. The **population** in which the study will be based.
2. The health determinant or **exposure** that will be studied.
3. The way in which the health **outcome** or **disease** will be defined and ascertained.
4. The **induction period** that is expected to elapse between the study exposure and its effect.
5. The **expected change in incidence** associated with exposure.
6. **Cofactors** (potential confounders) that may interfere with the interpretation of the study's results.
7. The plausible **causal mechanism** by which the exposure induces its effect
8. The **sample size** required to demonstrate an effect.

Illustrative Example 5.1 Elements of an epidemiologic hypothesis

Let us consider the above list of study design elements (1–8) through a hypothetical example. Use of combination oral contraceptives (COCs) is a known risk factor for venous thromboembolism (VTE). COCs are composed on an estrogen component and progestin component. Let us test the **hypothesis** that the estrogen dose in COCs is related to VTE risk. A study to address these factors could be set up as follows:

1. Population: Let us test this hypothesis in oral contraceptive users between the ages of 15 and 44 years of age in women enrolled in a particular health maintenance organization in the historical period 1985–1990.
2. Exposure: The exposure is ''high dose COCs.'' Let us define ''high dose'' as COCs that contain 50 μg of estrogen or more. The nonexposed group will be those women using ''low dose COCs'' containing 35 μg of estrogen or less.
3. Disease: The outcome of VTE will be diagnostically confirmed and treated venous thromboembolism, which includes deep venous thrombosis and pulmonary embolism.
4. Induction period: Effects are expected to be nearly immediate. Therefore, exposure will be defined as ''current use.''
5. Expected change in incidence: Reducing the dose from ''high dose'' to ''low dose'' is expected to cut the incidence of VTE in half from 8 per 10 000 person-years to 4 per 10 000 person-years.
6. Cofactors: Data on other risk factors for VTE such as age, recent surgery, trauma, the post-partum period, life-threatening illness, and other potential determinants will be collected and evaluated during the analysis of the data.
7. Mechanism: Exogenous estrogen is known to increase thrombogenesis and decreases fibrinolysis.
8. Sample size: A study of 60,000 women using high-dose formulations and 60,000 women using low-dose formulations followed for a year is expected to detect a twofold difference in risk with 80% power and an alpha level of 5%.

As data emerge from a study, and understandings increase from other sources, foreseeable facets of the study are adjusted. As new questions arise, additional hypotheses are articulated to guide further study.

Variables

Our goal is to make biobehavioral and contextual inferences about the effects of health determinants (''exposures'') on health outcomes (''diseases''). In generic terms we ask: ''Is there a causal relation between the explanatory exposure E and disease outcome D?''

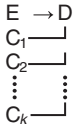
Exposure E → Disease D?

Exposure E in the study may represent any explanatory physiological factor, personal attribute, environmental exposure, social determinant, or medical intervention thought to influence health. This is sometimes referred to as the **independent variable** in a statistical analysis.

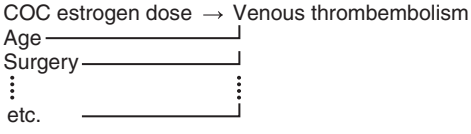
Disease D may represent any disease, illness, injury, response, or study endpoint. This is often referred to as the **dependent variable** in statistical analyses.

In addition to exposure E and health-outcome D, the study addresses other factors associated with exposure E and health outcome D. These additional factors are referred to as **potential confounders**, **control variables**, **extraneous variables**, or **cofactors**. Let us refer to co-variables as C_1, C_2, and so on.

The objective of the analytic epidemiologic research is to determine the effects E on D while accounting for the contributions of C_1, C_2, ... C_k.



For the hypothetical study addressed in Illustrative Example 5.1, these variables correspond to:



Data

The data that constitute variables E, D, C_1, C_2, ..., C_k are derived from a variety of sources. Examples of data sources include interviews with study subjects, self-administered questionnaires, employment records, environmental records, health care records, social services records, physical examination, examination of biological specimens, and various types of diagnostic tests. The type of data used in a particular study will depend on the research question being addressed, the cost of obtaining data, the need for confidentiality, the type population being studied, and the available technologies.

Information in medical records is abstracted and coded before analysis. Careful training of record abstractors is essential in order to obtain accurate and uniform information. It is advisable to blind interviewers and medical abstractors to the study hypothesis before data are collected. It is also necessary to obscure or remove sensitive

information from medical records in order to blind the abstractors to information that may influence objectivity when assigning codes and to protect the privacy of study subjects. This will prevent conscious and unconscious biases from entering into the abstraction process. When more than one medical record abstractor is involved in a study, it is wise to produce separate analyses for their data to check for consistency of results. Lack of consistency, such as a positive association derived by data from one reviewer but not the other, is cause for concern.

Data collection forms used by interviewers should be brief and simple. The art of asking clearly worded, non-ambiguous, and non-presumptuous questions takes extra thought and planning.[a] Data collection instruments must be piloted and tested to remove ambiguities and redundancies before being used in the actual study. Completed data forms should be reviewed by a study coordinator before being entered and validated in creating the data files that comprise the database for the study.

5.2 Ethical conduct of studies involving human subjects

Table 5.1 lists three ethical principles for conducting research using human subjects as specified in The Belmont Report (1979). These include respect for persons, beneficence, and justice. As part of the principle of respect for persons, study subjects must freely give their **informed consent** before participating in a study. This implies that the subjects are given a chance to ask questions, are not coerced, are under no obligation to participate in the study, and may withdraw from the study at any time. A signed statement of consent is required.

Ethical guidelines are safeguarded by human subjects committees known as **institutional review boards (IRBs)**. IRBs are committees composed of researchers, clinicians, administrators, and laypeople who review the study protocol before the study is begun. Their primary objective is to ensure the ethical treatment of human subjects and to oversee informed consent procedures.

Because experiments involve treatments and interventions, additional ethical constraints are required. To ethically assign treatments, none of the treatments can be known to be superior to any other. Treatments that present special hazards cannot

Table 5.1 Principles of ethical research involving human subjects.

Ethical principle	Application
Respect for persons	Informed consent given freely. Implies ability to comprehend consequences; confidentiality is maintained for private information.
Beneficence	Risk and benefits are assessed. Benefits can be direct, indirect, collateral, or aspirational. Harms may involve physiological, psychological, or socioeconomic consequences.
Justice	Selection of study groups should be inclusive, equitable, and avoid exploitation.

Source: The Belmont Report (1979).

[a] See the classical handbook *The Art of Asking Questions* by S.L. Payne, Princeton University Press, 1951.

be ethically assigned and, just as importantly, treatments that are believed to be beneficial cannot be ethically withheld. Therefore, a true state of uncertainty or ''balanced doubt'' about the pros and cons of the intervention must exist before it can be submitted to a trial. This balanced void of knowledge is referred to as **equipoise**.

Separate from the IRB, studies involving interventions require a **Data and Safety Monitoring Board** (DSMB). The DSMB is an independent group of outside experts that periodically reviews and evaluates accumulated evidence from the study to monitor its safety and progress. The job of the DSMB is to make recommendations concerning the continuation, modification, or termination of the study.

5.3 Selected study design elements

Let us now address five important elements of epidemiologic study design. These elements are: (a) necessity of including a referent (''control'') group, (b) the distinction between experimental and non-experimental (observational) studies, (c) the unit of observation, (d) the difference between cross-sectional and longitudinal observations, and (e) cohort and case–control samples.

Necessity of a referent (''control'') group

Under all but exceptional circumstances, the only way to demonstrate whether a given exposure is the cause of an outcome is to compare groups to see if different combinations of exposures explain variations in the outcome under a variety of circumstances. Thus, etiologic studies require at least two groups. One group, the **index group**, is exposed to the factor thought to influence occurrence of the study outcome. The other group, the **referent** or **control group**, remains unexposed to provide a reference for comparison.

The effects of the exposure cannot be judged without the benefit of the referent group. Consider an experiment to determine the effectiveness of a treatment. If a certain number of patients recover following treatment, how would we know whether recoveries were due to the treatment or whether recoveries merely represented the spontaneous recovery rate or perhaps some unmeasured factor? Might the treatment have delayed recovery? On the other hand, by administering the treatment to one group of patients while leaving a similar group of patients untreated, observed difference in recovery could then be ascribed to either the treatment, some unnoticed difference in the groups, or to chance. Without the baseline of observation provided by the referent group, it is impossible to determine whether the exposure had a positive effect, a negative effect, or no effect at all.

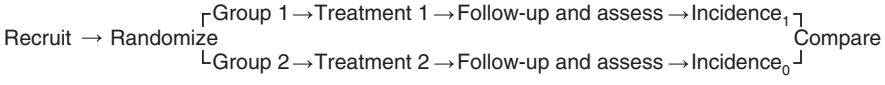
This same line of reasoning applies to studies that address natural exposures. If a person develops a brain tumor and is a frequent cell phone user, or lives near a toxic waste dump, or is exposed to whatever the media has identified as the hazard of the moment, we might be tempted to attribute their condition to any one of these exposures. Nonetheless, the outcome, however unfortunate, cannot be attributed to any particular exposure without further scrutiny. The question of course is not whether there is an association in the minds of any particular individual. The question is whether the specific exposure contributed to the causal mechanism behind the

brain tumor. If the exposure is causal, we would expect an increased occurrence of the outcome in those exposed to the causal factor relative to those who are not exposed, all other things being equal.

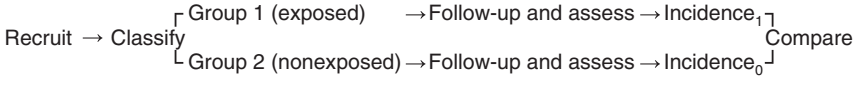
Experimental versus observational study designs

The primary way to classify comparative studies in epidemiology is as either experimental or non-experimental (observational). In **experimental studies ("trials")**, the investigator introduces or withholds an exposure in order to observe its effects. The experimental allocation of the study exposure can be based on chance mechanisms (**randomized trials**) or on other mechanisms built into the study's protocol (**nonrandomized trials**). Randomized designs are superior to nonrandomized trials for reasons that will soon become evident.

In a simple randomized controlled trial, individuals are randomly assigned to either a treatment group or a control group. The treatment group receives the experimental intervention. The control group receives either an inert intervention (placebo) or an alterative active intervention. Study subjects are then followed over time to assess study outcomes.



Because of practical and ethical concerns, however, opportunities for experiments using human subjects are often limited. Thus, most epidemiologic studies are non-experimental. **Non-experimental epidemiologic studies** are often referred to as **observational studies**. In contrast to experimental studies, observational studies do not assign treatments to study participants. Instead, subjects are studied under natural circumstances that are thought to be revealing. In a simple observational cohort design, subjects are *classified* as either ''exposed'' or ''nonexposed'' to the study factor of interest and are then followed and assessed for the study outcome. Incidences in the two groups are then compared.[b]



Illustrative Example 5.2 Women's Health Initiative (WHI)

The Women's Health Initiative (WHI) was a major 15-year research initiative sponsored by the National Heart, Lung, and Blood Institute. The objective of this program was to address the common causes of death, disability, and poor quality of life in postmenopausal women, with special emphasis on cardiovascular disease, cancer, and osteoporosis. This program included both experimental and observational studies.

Experimental elements of the WHI project were designed to test the effects of postmenopausal hormone therapy, diet modification, and calcium and vitamin D supplements on heart disease,

[b] This describes a prospective observational cohort study.

fractures, and breast and colorectal cancer. In the hormone trial, for example, participants were randomly assigned to groups that received either a pill containing estrogen plus progesterone or an identical-looking pill that contained no active ingredients. Incidences of various health outcomes (e.g., coronary disease) were monitored over time in the study subjects.

Observational studies in the WHI program complemented the experimental studies by providing estimates of the extent to which various risk factors predicted heart disease, cancers, fractures, and other adverse health outcomes. Observational studies in the WHI tracked the experience of 93 676 postmenopausal women between the ages of 50 and 79. Women who joined the observational study were not required to take any medication or change their health habits while being monitored.

One of the leading concerns in both experimental studies and observational studies concerns the ''fairness'' of comparison. For comparisons to be meaningful, the groups should be similar in all relevant ways *except* for the exposure being studied. Randomization helps achieve ''like-to-like'' comparison by balancing the distribution of measured and unmeasured extraneous factors that could otherwise confound the interpretation of results. In contrast, non-experimental observational studies must rely on other methods to address group comparability. This important issue will be revisited throughout this book. Let us start our consideration of confounding with this example concerning the effects of hormone replacement therapy in menopausal women.

Illustrative Example 5.3 Hormone replacement therapy

It was not too long ago that women routinely used hormone replacement therapy (estrogen plus progestin) at around the time they reached menopause. One of the reasons for the wide acceptance of hormone replacement therapy was that observational studies had shown women who took hormones at menopause had a lower risk of cardiovascular disease than women who did not. However, in 1998, a randomized trial of hormone therapy in women who had already had heart disease found no benefit to hormone use (Hulley *et al*., 1998). In 2002, the WHI estrogen plus progestin experimental trial in healthy women (see Illustrative Example 5.2) showed that indiscriminant use of hormone replacement therapy actually increased the risk of cardiovascular diseases. Thus, the randomized trials contradicted the results of the earlier observational studies. Upon further scrutiny, most epidemiologists now believe that the results of the earlier non-experimental (observational) studies could be ascribed to the fact that women who tended to use hormone replacement therapy postmenopausally tended to be healthier and of higher socioeconomic status than women who did not seek this type of treatment. The experimental studies avoided this problem by assignment of the exposure by chance mechanisms, averting self-selection of hormone replacement.

Unit of observation

The unit of observation in an epidemiologic study refers to the level of aggregation upon which measurements are available. This can vary from person-level data to region-level data, and everything in between:

Persons ↔ Families ↔ Social groups ↔ Neighborhoods ↔ Regions ↔ Nations

Consider, for example, studying the effects of smoking. We can measure the smoking habits of individuals in terms of, say, ''the number of cigarettes an individual smokes per day.'' In contrast, we can measure a *region's* in terms of ''per capita cigarette

consumption.'' The former is an individual-level unit of observation, while the latter is an aggregate-level unit of observation.

Because pathophysiological and behavioral phenomena occur at the individual level, most epidemiologic studies use individual-level data. However, studies that address social and environmental phenomena may combine individual-level measures with environmental- and aggregate-level measures in what is known as a multilevel analyses, as discussed in Section 4.3.

Longitudinal versus cross-sectional observations

Person-level observations may be longitudinal or cross-sectional. **Longitudinal** observations address individual experiences over time. In contrast, **cross-sectional** observations do not permit the accurate time-sequencing of events within individuals.

Note that the study design feature that distinguishes longitudinal observations from cross-sectional observations is the ability to accurately place the events *for individuals* on a time-line. This is not to be confused with prospective and retrospective data, which addresses the proximity of events in time to the time of data collection (see Section 7.4). Note that, longitudinal observations address individual experience over time. This is not to be confused with serial cross-sectional rates that are tracked over time (discussed further in Section 7.7).

A single serological ascertainment for HIV, for example, is cross-sectional even if collected prospectively, because it is unable to determine *when* an individual became seropositive. To derive longitudinal data for HIV status, one would need a multiple of serological measurements to be obtained over time starting with seronegative individuals. These longitudinal measurements could then derive the approximate dates of seroconversion for individuals.

Longitudinal data are preferable when conducting etiologic research, because for a factor to be causal, it must clearly precede the event it caused by a reasonable amount of time. However, for characteristics that do not change over time (e.g., genetic factors), it matters little whether the measurement is longitudinal or cross-sectional, because if we know the status of this factor now, we also know its status in the past. In addition, many human habits that are potentially changeable, such as dietary choices, display some degree of long-term permanence. For stable characteristics such as these, the current status of the attribute serves as a suitable proxy for its longitudinal equivalent.

Cross-sectional epidemiologic research was particularly common in the early and mid-20th century, especially in the period around World War II. An example follows.

Illustrative Example 5.4 Mental health survey (cross-sectional study)

Table 5.2 contains data from an historically important example of a community mental health survey. Data were compiled by a team of psychiatrists and sociologists using data from a New Haven, CT, USA, urban community hospital. Social economic status (SES) (the exposure) was based on a combination of factors about neighborhood of residence, occupation, and education. Data demonstrate a positive association between low socioeconomic status and psychosis, while showing a negative association between low socioeconomic status and neurosis.

Table 5.2 Prevalence per 100 000 of psychosis and neurosis by social economic status, CT, USA 1950.

Socio-economic status	Psychosis	Neurosis
1 and 2 (high)	188	349
3	291	250
4	518	114
5 (low)	1505	97

[a]Prevalences have been adjusted for sex and age.
Source: Hollingshead and Redlich, (1953; 1964, pp. 230, 235).

With the increasingly recognized importance of chronic diseases with long induction periods, weaknesses with cross-sectional associations soon became evident. Example of these weaknesses that are evident in the above Illustrative Example include:

- **Detection bias**. Some persons with mental disorders may not come to the attention of the health-care system. These differences in detection may partially account for the observed associations. Psychosis, for example, may come to the attention of psychiatrists through legal intervention. Since social class is strongly correlated with legal interventions, the selective forces that bring patients to care are also strongly correlated with the study exposure and study outcome. This could exaggerate or even create the positive association between low SES and psychosis.
- **Diagnostic bias**. There may be diagnostic preferences associated with both neurosis and psychosis. For example, state hospitals, where the lower SES patients are more likely to be seen, may more likely diagnose psychosis, whereas private providers may be more likely to diagnose neurosis in the same case. Thus, the observed rates may be a spurious function of the provider type.
- **Reverse-causality bias**. Another source of bias that must be considered is called ''reverse-causality bias'' or ''cart-before-the-horse bias.'' Because data are cross-sectional, it is difficult to establish the correct temporal sequence of events: the time order of the exposure–disease may be turned around. An essential property of a causal factor is for the exposure to precede the onset of disease. With reverse causality bias, the temporal sequencing of events is reversed. Although one may be tempted to say that low social status causes psychosis, another plausible explanation is that psychosis causes downward social mobility (because psychotics cannot always maintain the normal social relations required to maintain a reasonable level of income, for instance). With cross-sectional data, the proper temporal sequence can only be assumed.
- **Incidence–prevalence bias**. Varying duration of illness may confuse the interpretation of results in a type of bias called ''prevalence–incidence bias'' (Neyman, 1955). Prevalence is related to both incidence and average duration of illness (see Chapter 3, Section 4). In studying the prevalence of a condition, therefore, cases of long duration are more heavily weighted than those of short duration. If the socioeconomic groups had similar incidences of neurosis, for example, but SES groups had more persistent diagnoses, an apparent gradient in prevalence would exist with no difference in incidence. Indeed, this is what the researchers found in a subsequent analysis (Table 5.3) in which the incidence of neurosis was not linked

Table 5.3 Incidence, reentry, continuous, and prevalence per 100 000 of neurosis and psychosis by socioeconomic status.[a].

SES	Incidence	Reentry	Continuous	Prevalence
Neurosis:				
1 and 2 (high)	69	44	251	349
3	78	30	137	250
4	52	17	82	114
5 (low)	66	35	65	97
Psychosis:				
1 and 2 (high)	28	44	117	188
3	36	38	217	291
4	37	42	436	518
5 (low)	73	88	1344	1505

[a]Data have been adjusted for sex and age.
Source: Hollingshead and Redlich, (1964, p. 235).

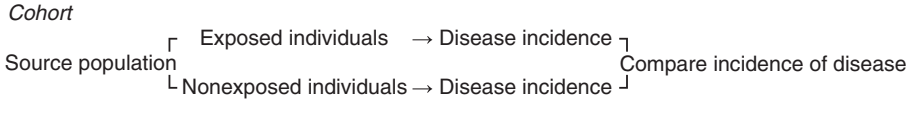
to socioeconomic status (Table 5.3). In contrast, the positive association between low socioeconomic status and psychosis persisted when the analysis was restricted to incident cases.

Historical efforts to correct the weaknesses of cross-sectional surveys lead to greater rigor in epidemiologic methods and to the development of modern cohort and case–control methods.

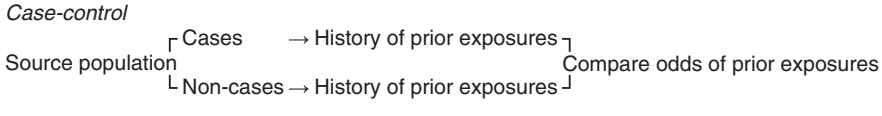
Cohort versus case–control samples

Longitudinal studies track health-related events and experiences in individuals over time. The two primary types of observational studies that permit this type of tracking are cohort studies and case–control studies.

Cohort studies begin by identifying disease-free individuals. Study subjects are then classified according to risk factors thought to be associated with future disease occurrence. A period ensues during which disease is monitored. Incidences of events are then tallied and compared among the exposure groups.



In contrast to cohort studies, **case–control studies** begin by identifying people with the disease being studied (the **case series**). They then select non-cases from the same population that gave rise to the cases (the **control series**). Exposures to risk factors thought to be predictive of the study outcome are then ascertained retrospectively in cases and controls.



Note that the key distinction between cohort and case–control studies is the way in which subjects are sampled for study: cohort studies begin with disease-free study subjects, while case–control studies begin with diseased (cases) and disease-free (controls) study subjects. Nevertheless, both cohort and case–control studies rely on the reconstruction of events in individuals over time and are thus longitudinal.[c] Examples of cohort and case–control studies follow.

Illustrative Example 5.5 Oral contraceptive estrogen dose and venous thromboembolism (cohort study)

A cohort study examined the incidence of venous thromboembolism in 234 218 women using oral contraceptives with varying amounts of estrogen (Gerstman *et al*., 1991). The study was restricted to women taking combination oral contraceptives. Women taking formulations containing less than 50 μg of estrogen were classified as ''low-dose users.'' Women taking formulation containing exactly 50 μg of estrogen were classified as ''intermediate-dose users.'' Women taking formulations containing more than 50 μg of estrogen were classified as ''high-dose users.'' The experience of each study subject was tracked for the occurrence of venous thromboembolism (VTE). The rates of VTE in each group were

- Low-dose users: $\dfrac{53 \text{ cases}}{12.7 \times 10^4 \text{ person-years}} = 4.2 \text{ per } 10\,000 \text{ person-years}$
- Intermediate-dose users: $\dfrac{69 \text{ cases}}{9.8 \times 10^4 \text{ person-years}} = 7.0 \text{ per } 10\,000 \text{ person-years}$
- High-doses users: $\dfrac{20 \text{ cases}}{2.0 \times 10^4 \text{ person-years}} = 10.0 \text{ per } 10\,000 \text{ person-years}$

Thus, progressively higher estrogen doses were associated with VTE rates in this cohort.

Illustrative Example 5.6 Toxic shock syndrome (case–control study)

Toxic shock syndrome (TSS) is an illness characterized by high fever, vomiting, diarrhea, confusion, and an exfoliating skin rash. It is fatal in 3–15% of cases and is caused by the exotoxin of a particular strain of *Staphylococcus*. In late 1979 and early 1980, the Centers for Disease Control received an unusual number of reports of TSS from state health departments in Wisconsin, Minnesota, Illinois, Utah, and Idaho (CDC, 1980a; 1980b). The cases occurred almost exclusively in women of childbearing age. Several case–control studies were completed in the wake of these reports. One of the case–control studies evaluated prior exposures to risk factors in 52 cases and 52 age- and sex-matched controls. All of the 52 cases in this study had used tampons during their menstrual periods coincident with the onset of illness. In contrast, 44 (85%) of the 52 control study subjects had used tampons during their prior menstrual period. Thus, cases were more likely to be tampon users than controls (Shands *et al*., 1980). In this same study, among the 44 case–control pairs in which both study subjects had used tampons, 42 (95%) of the 44 cases used tampons continuously throughout menstruation. In contrast, 34 (77%) of 44 controls did similarly. Thus, among the tampon users, cases were more likely to use tampons continuously. This information provided clues that led eventually to the discovery of a highly absorbent brand of tampon as a risk factor for the proliferation of toxogenic *Staphylococci* as the cause of TSS.

Notice the relatively small number of study subjects in the case–control study in Illustrative Example 5.6 (52 cases and 52 controls; 104 subjects total). Compare this to the cohort study in Illustrative Example 5.5 (234 218 study subjects). Although

[c] It is incorrect to think of cohort studies as strictly ''prospective'' and case–control studies as ''retrospective,'' as will be demonstrated in Chapters 7 and 8.

the case–control study was small, it was still able to provide a reliable estimate of the relation between the antecedent exposure (tampon use) and subsequent disease (toxic shock syndrome).[d] This type of efficiency is one of the key advantages of the case–control method. Cohort studies and case–control studies will be covered in greater detail in Chapters 7 and 8, respectively.

5.4 Common types of epidemiologic studies

The aforementioned study design elements define the main types of studies encountered in epidemiology. Studies are initially classified as either observational (non-experimental) or experimental. Observational studies are divided according to whether their unit of observation is at the aggregate-level (**ecological studies**) or personal-level. Person-level studies are divided according to whether they are **cross-sectional** or longitudinal. Finally, longitudinal studies are classified as either **cohort** or **case–control**.

Experimental studies are divided into **community trials**, which address intervention applied at the group level (e.g., public health information campaigns), **field trials**, which are primary prevention interventions applied to individuals (e.g., vaccine trials), and **clinical trials**, which address therapeutic interventions of individuals in the treatment of illness (e.g., chemotherapy trials). Figure 5.1 presents a schematic of this classification scheme to help discern the common types of epidemiologic studies.

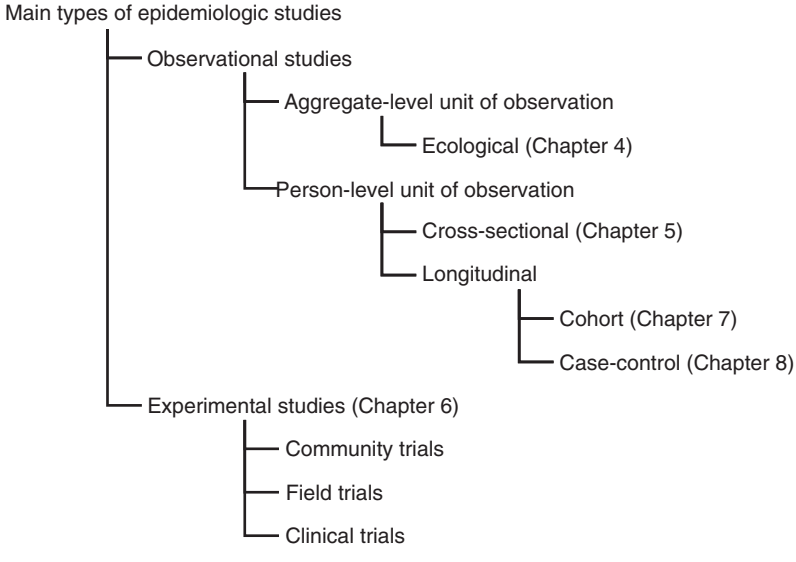


Figure 5.1 Common types of epidemiologic studies and where to find them covered in this text.

[d] In Chapter 8 we will learn how to quantify relationships from case–control studies with a statistic known as the odds ratio.

Exercises

5.1 Study types. For each of the brief descriptions below, identify the study's exposure and disease. In addition, determine whether the study is experimental, ecological, cross-sectional, cohort, or case–control. Briefly, justify your response.

(A) Epidemiologists suspect that avian adeno-associated virus is caused by exposure to poultry. Serum samples from poultry workers and the general population are tested to determine the proportion of individuals positive for avian A–V antibody in each group.

(B) The behavioral pattern identified as Type A behavior is characterized by a hard-driving personality susceptible to anger and time urgency. This type of behavior is thought to be associated with increased risk for coronary heart disease. Type A behavior is ascertained in a group of men in a postcoronary disease rehabilitation program. Men not falling into the Type A category are classified as Type B. Type A and Type B men are then followed for 5 years to assess for the recurrence of acute coronary symptoms.

(C) Investigators studying bus company employees want to test the hypothesis that occupational stress causes high blood pressure. Two groups of employees are compared: bus drivers and office workers for bus companies in the same salary range as bus drivers matched on age, sex, race, and length of employment. The investigators take blood pressure measurements of all study subjects and find that the mean blood pressure of bus drivers is higher than that of office workers.

(D) One hundred newly diagnosed breast cancer patients are interviewed to determine dietary histories. A similar number of healthy first-degree relatives (mothers or sisters) are interviewed in a similar manner. We compare the proportion of women reporting a history of high dietary fat consumption in the two groups.

(E) Fifteen hundred men working for an aircraft manufacturing company are recruited to participate in a study of coronary heart disease. Every 3 years, study subjects are examined for the onset of disease. Coronary disease rates are compared among groups defined by various personal characteristics (e.g., job category, blood pressure, diet type, exercise program) that were recorded at the beginning of the study.

(F) A sample of sedentary middle-aged men is selected from four census tracts. Each man is examined for coronary heart disease. Subjects having evidence of pre-existing coronary heart disease are excluded from further study. Eligible subjects are assigned to either a group that is coached to pursue regular moderate exercise or a control group that gets a sham intervention. Subjects are examined semiannually to determine their cardiac health.

(G) One hundred incident cases of infectious hepatitis and 100 healthy neighbor controls are asked about their history of eating raw clams and oysters over the preceding year.

(H) Questionnaires are mailed to every tenth person listed in a city directory. Each person is asked to list his or her age, sex, occupation, socioeconomic status, smoking habits, and musculoskeletal symptoms during the preceding 7 days. About three-quarters of the questionnaires are completed and returned. The

frequency of various types of musculoskeletal symptoms is compared in smokers and nonsmokers, controlling for age, sex, and socioeconomic status.

(I) An investigator collects information on the size of manufacturing plants and their rates of accidents. She finds that the five largest plants have accident rates that are 50% higher than the five smallest plants.

5.2 Driving while talking. You have developed a hypothesis that automobile drivers that talk on their cell phones have higher rates of fatal automobile accidents than those who do not, and want to test this hypothesis with a cohort study.

(A) What is the exposure being studied by this hypothesis? How might you measure the exposure in your cohort study?

(B) How might you identify cases in this study?

(C) What additional factors would you strive to measure in this study?

(D) What difficulties might be encountered when measuring driving characteristics?

(E) Since fatal automobile accidents are, fortunately, a rare occurrence, we have reconsidered pursuing this hypothesis with a cohort study and are now considering a case–control study. How would you design a case–control study to test the hypothesis?

5.3 Agricultural injuries. A cohort study evaluated risk factors for agriculture-related injuries in African–American and Caucasian farmers, and African–American farm workers (McGwin *et al.*, 2000). A total of 1 246 subjects (685 Caucasian owners, 321 African–American owners, and 240 African–American workers) were enrolled between January 1994 and June 1996. Demographic, farming, and behavioral information was collected at baseline. Subjects were contacted biannually to monitor the occurrence of any agriculture-related injury. Data from this study are presented in this table:

Group	Agricultural-related injuries	Person-years of observation
Caucasian farm owners	67	2047
Af–American farm owners	27	821
Af–American workers	37	359

(A) List the two exposure variables addressed by this study.

(B) Identify the study outcome.

(C) Explain why experimentation was not possible when addressing this issue.

5.4 Open population rates. Explain why rates derived in open-populations are *not* longitudinal.

5.5 Classify the study. After reading each of the following passages, identify the exposure variable, primary study outcome variable, and any extraneous variables identified in the question. Then classify each study as either a case report or

case series, ecological study, cross-sectional survey, observational cohort study, case–control study, or epidemiologic experiment.

(A) A 39-year old woman presents with a mild sore throat, fever, malaise, and headache and is treated with penicillin, for presumed streptococcal infection. She returns in a week with hypertension, fever, rash, and abdominal pain. She responds favorably to chloramphenicol, after a diagnosis of Rocky Mountain spotted fever is made.

(B) 50 patients with thyroid cancer are identified and surveyed by patient interviews to identify prior exposure to radiation.

(C) Patients admitted for carcinoma of the stomach and patients without a diagnosis of cancer are interviewed about their chewing tobacco history to assess the possible association of chewing tobacco and gastric cancer.

(D) Data on median income for households in census tracts within a large metropolitan county in the United States were obtained from the Census Bureau's Current Population Survey. Air pollution levels were measured in these same census tracts during a period of one month. The data were analyzed using a geographic information system (GIS) to produce maps showing pollution and income levels by census tract.

(E) In a large study carried out in the United Kingdom, the death-rate from diseases of the circulatory system in women who had used oral contraceptives was five times that of women who had never used oral contraceptives.

(F) A nutritionist has developed the hypothesis that providing breakfast in elementary schools will decrease obesity among the students. Eight elementary schools agree to participate in the study, four of which will be ''treatment schools'' and four of which will receive no special intervention. The assignment of the treatment will be by randomization. The heights and weights of students will be monitored over the next 3 years.

Review questions

R.5.1 What is the primary distinction between experimental and observational study designs?

R.5.2 What is the primary purpose of randomizing the study exposure in a randomized experiment?

R.5.3 Is ''current health status'' a longitudinal or cross-sectional variable?

R.5.4 What makes a study longitudinal as opposed to cross-sectional?

R.5.5 What distinguishes cohort studies from case–control studies?

R.5.6 What makes a study ecological?

R.5.7 Match each study design with its brief description.
Designs: experimental, case–control, cohort, cross-sectional, ecological.
Brief descriptions:
1 Based on aggregate-level data.
2 Based on non-longitudinal data on individuals.
3 Diseased and nondiseased individuals compared with respect to prior exposures.

4 Exposed and nonexposed individuals compared with respect to incidence of outcomes.
5 Exposures assigned to study subjects (usually randomly) as part of the study's protocol.

R.5.8 True or false? Whether a measurement is longitudinal or cross-sectional depends on when data are collected in relation to dates of actual occurrence. Explain your response.

R.5.9 Match the bias with its description.
Bias: detection bias, reverse-causality bias, prevalence-incidence bias.
Description:
1 Long-duration cases contribute to study outcomes to a greater extent than short-duration cases.
2 The exposure increases likelihood of diagnosis but does not cause the ailment.
3 The ''disease'' causes the ''exposure,'' not vice versa.

R.5.10 What is equipoise?

R.5.11 What is informed consent?

R.5.12 List the three guiding principles of the Belmont Report.

R.5.13 What does IRB stand for?

R.5.14 What do IRBs do?

References

Belmont Report. (1979). The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Retrieved 18 September 2012, from *http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html*.

CDC (1980a) Toxic shock syndrome – United States, *MMWR*, **29** (May 23, 1980), 229–230.

CDC (1980b) Toxic shock syndrome – United States, *MMWR*, **29** (June 27, 1980), 297–299.

Gerstman, B.B., Piper, J.M., Tomita, D.K., Ferguson, W.J., Stadel, B.V., and Lundin, F.E. (1991). Oral contraceptive estrogen dose and the risk of deep venous thromboembolic disease. *American Journal of Epidemiology*, **133** (1), 32–37.

Hollingshead, A.B., and Redlich, F.C. (1953). Social stratification and psychiatric disorders. *American Sociological Review*, **18**, 163–169.

Hollingshead, A.B., and Redlich, F.C. (1964). *Social Class and Mental Illness: A Community Study*, John Wiley & Sons, Inc., New York.

Hulley, S., Grady, D., Bush, T., Furberg, C., Herrington, D., Riggs, B., *et al*. (1998). Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *Journal of the American Medical Association*, **280** (7), 605–613.

McGwin, G., Jr.,, Enochs, R., and Roseman, J.M. (2000). Increased risk of agricultural injury among African-American farm workers from Alabama and Mississippi. *American Journal of Epidemiology*, **152** (7), 640–650.

Neyman, J. (1955). Statistics—servant of all sciences. *Science*, **122**, 401–406.

Shands, K.N., Schmid, G.P., Dan, B.B., Blum, D., Guidotti, R.J., Hargrett, N.T., *et al*. (1980). Toxic-shock syndrome in menstruating women: association with tampon use and Staphylococcus aureus and clinical features in 52 cases. *New England Journal of Medicine*, **303** (25), 1436–1442.

CHAPTER 6
Experimental Studies

6.1 Introduction
6.2 Historical perspective
- Comment regarding use of the term ''natural experiment''

6.3 General concepts
- The control group
- Randomization and comparability
- Checking group comparability
- Recruitment and eligibility criteria
- Follow-up and outcome ascertainment
- Intention-to-treat analysis versus per-protocol analysis

6.4 Data analysis
- Measures of effect
- Statistical inference
- Sample size requirements

Exercises
Review questions
References

6.1 Introduction

Epidemiologic experiments are also called **trials**. The word *trial* comes from the Anglo-French root *trier*, meaning ''to try'' or ''to put something to a test.'' Epidemiologic trials put either preventive or therapeutic measures to the test. The three main types of trials in epidemiology are:

- **Field trials**, which are used to address the efficacy of preventive interventions applied to individuals (e.g., a vaccine trial).
- **Community trials**, which are used to address the efficacy of preventive interventions applied at the group level (e.g., a health education campaign).
- **Clinical trials**, which are used to address the efficacy of therapeutic interventions in ill individuals (e.g., a chemotherapy trial in the treatment of cancer).

Let us start by considering of a couple of examples.

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

Illustrative Example 6.1 WHI estrogen plus progesterone trial

The Women's Health Initiative (WHI) included a large randomized controlled primary prevention trial that addressed whether women in the decades of life following menopause should take estrogen plus progesterone in order to prevent chronic diseases (WHI, 2002). The study included 16 608 postmenopausal women aged 50–79 years who were randomized to form two groups. Group 1 received conjugated estrogens plus progesterone in tablet form (n = 8506). Group 2 received an identical looking placebo tablet (n = 8102). The two main disease outcomes in this trial were coronary heart disease and invasive breast cancer. After an average of 5.2 years of follow-up, the trial was stopped because the risks of continued use of the active treatment exceeded its benefits. For example, the incidence proportion of fatal and nonfatal coronary events in the estrogen plus progesterone group was $\frac{164}{8506}$ = 0.019 28 or 19.3 per 1000. In the placebo group, the incidence was $\frac{122}{8102}$ = 0.015 06 or 15.1 per 1000. Figure 6.1 plots the occurrence of coronary events in the groups over time. This plot demonstrates that differences between the groups became evident soon after randomization and persisted through the follow-up period.

Whereas the prior example illustrated a field trial in which individuals were randomized into a treatment or control group, other trials randomized the study intervention on a group-by-group basis, as illustrated in this example.

Illustrative Example 6.2 Vitamin A trial in Sumatra

This study sought to determine if childhood mortality due to high rates of respiratory disease and diarrhea could be reduced by vitamin A supplementation; 450 villages in Sumatra were randomly assigned to either participate in a vitamin A supplementation program (n = 229) or to serve as a control village (n = 221) (Sommer *et al*., 1986). Vitamin A capsules were distributed to preschool children 1–3 months after enrollment in the treated villages and again 6 months later. The one-year mortality rate in children (1–5 years) in the villages that received vitamin A supplementation was $\frac{53 \text{ deaths}}{10\,919 \text{ children}}$ = 4.9 per 1000. The mortality rate in the control villages was $\frac{75 \text{ deaths}}{10\,231 \text{ children}}$ = 7.3 per 1000. Thus, the rate difference mortality was (4.9 per 1000) – (7.3 per 1000) = −2.4 per 1000, indicating that vitamin A had reduced childhood mortality substantially.

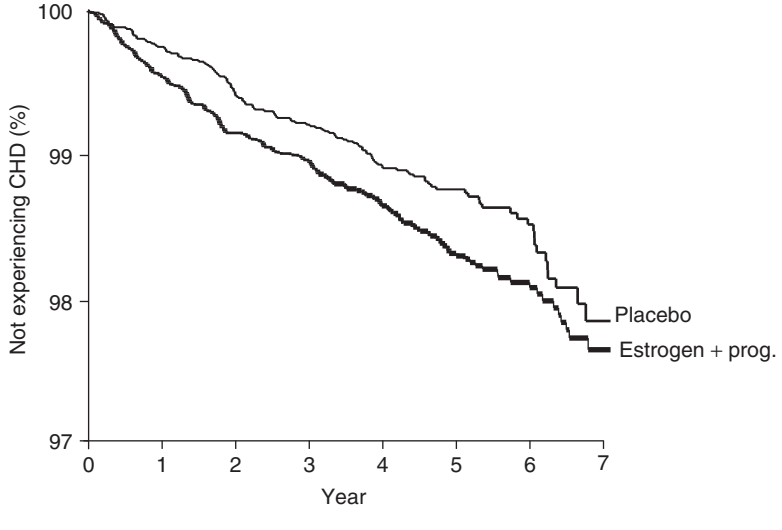


Figure 6.1 Percent not experiencing coronary disease, WHI estrogen plus progesterone trial (WHI, 2002).

Before delving further into modern epidemiologic trials, let us gain some historical context.

6.2 Historical perspective

The *idea* of an experiment in human health is really quite old. The earliest recorded account of a trial appears in the first chapter of the Book of Daniel in the *Old Testament* (Lilienfeld, 1982). In this story, Daniel requests a 10-day trial comparing a diet of the ''king's food'' with a standard diet of leguminous plants. Daniel predicts superior results on the standard diet of leguminous plants, which he requests for his people. After 10 days on the respective diets, Daniel recommends:

> Then let our countenances be looked upon before thee, and the countenances of the youths that eat of the king's food.

Apparently, the proposal was accepted, since the story goes on to tell:

> So, he hearkened unto them and tried them in this matter, and tried them for ten days . . .

The results favored the diet of leguminous plants:

> . . . at the end of the ten days [the group eating the diet of leguminous plants had] countenances [which] appeared fairer and they were fatter in the flesh, than all the youths that did eat the king's food.
>
> *Biblical quotes cited in Lilienfeld*, 1982, p. 4

One of the earliest descriptions of a randomized trial was provided by the Belgian medicinal chemist **van Helmont** in 1662. Wishing to replace the theory-based approach of his peers with a more empirical approach, van Helmont wrote:

> Let us take out of the hospitals, out of the Camps, or from elsewhere, 200, or 500 poor People, that have Fevers, Pleurisies, &c. Let us divide them into halfes, let us cast lots, that one half of them may fall to my share, and the others to yours; . . . we shall see how many funerals both of us shall have: *But* let the reward of the contention or wager, be 300 florens, deposited on both sides.
>
> *Armitage*, 1983, p. 328

This passage describes a **randomized controlled trial** (**RCT**). It was **randomized** because the treatment was assigned to study subjects by mechanism based on chance (''let us cast lots''). It was **controlled** because there is a treatment and control group (''that one half of them may fall to my share, and the other to yours''). It was a clinical trial because it tested the efficacy of a treatment in the caring for the ill (''People, that have Fevers, Pleurisies, &c.''). Thus, the general idea of an RCT dates back many centuries.

A well-known historical example of a **nonrandomized trial** is that of James Lind's 1753 trial of treatments for scurvy. Lind assembled six elixirs and concoctions that he thought to be the most likely cures for scurvy. He then assigned a pair of scurvy-ridden sailors to each of the six treatments. Lind (1753) remarks, ''the most sudden and visible good effects were perceived from the use of the oranges and lemons.''

Hence the practice of supplying British sailors with citrus at sea and their derived nickname ''limeys.'' Box 6.1 presents Lind's experiment in his own words.

Box 6.1: Lind's trial (1753) in his own words

On the 20th May, 1747, I took twelve patients in the scurvy on board the *Salisbury* at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of their knees. They lay together in one place, being a proper apartment for the sick in the forehold; and had one diet in common to all, viz., water gruel sweetened with sugar in the morning; fresh mutton broth often times for dinner; at other times puddings, boiled biscuit with sugar etc.; and for supper barley, raisins, rice and currants, sago and wine, or the like. Two of these were ordered each a quart of cyder a day. Two others took twenty five gutts of elixir vitriol three times a day upon an empty stomach, using a gargle strongly acidulated with it for their mouths. Two others took two spoonfuls of vinegar three times a day upon an empty stomach, having their gruels and their other food well acidulated with it, as also the gargle for the mouth. Two of the worst patients, with the tendons in the ham rigid (a symptom none the rest had) were put under a course of sea water. Of this they drank half a pint every day and sometimes more or less as it operated by way of gentle physic. Two others had each two oranges and one lemon given them every day. These they eat with greediness at different times upon an empty stomach. They continued but six days under this course, having consumed the quantity that could be spared. The two remaining patients took the bigness of a nutmeg three times a day of an electuray recommended by an hospital surgeon made of garlic, mustard seed, *rad. raphan*., balsam of Peru and gum myrrh, using for common drink narley water well acidulated with tamarinds, by a decoction of which, with the addition of *cremor tartar*, they were gently purged three or four times during the course.

The consequence was that the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them being at the end of six days fit four duty. The spots were not indeed at that time quite off his body, nor his gums sound; but without any other medicine than a gargarism or elixir of vitriol he became quite healthy before we came into Plymouth, which was on the 16th June. The other was the best recovered of any in his condition, and being now deemed pretty well was appointed nurse to the rest of the sick . . .

As I shall have occasion elsewhere to take notice of the effects of other medicines in this disease, I shall here only observe that the result of all my experiments was that oranges and lemons were the most effectual remedies for this distemper at sea.

Source: *Lind* (1753).

Comment regarding use of the term ''natural experiment''

The term **natural experiment** has historically been used to refer to a study with a natural but fortuitous distribution of treatments that mimics an experiment. An early example of one such study was described by the barber-surgeon **Ambroise Paré** (circa 1510–1590). During the battle for the castle of Villaine in 1537, Paré ran out of the standard treatment for battle wounds which, at that time, was to douse the wound with boiling oil. Having run out of the standard treatment, Paré resorted to treat the wounds with a much less noxious treatment of ''digestive medicament.'' After the battle, Paré noted superior results with the alternative innocuous digestive medicament, stating:

> I raised myself very early to visit them, when beyond my hope I found those to whom I had applied the digestive medicament, feeling but little pain, their wounds neither swollen nor inflamed, and having slept through the night. The others whom I had applied the boiling oil were feverish with

> much pain and swelling about their wounds. Then I determined never again to burn thus so cruelly the poor wounded.
>
> Armitage and Colton, 1998, pp. 1–2

Although this type of observation has been historically referred to as a natural experiment, it is actually nonexperimental (observational) in nature, since use of the alternative treatment was not allocated as part of the study protocol. Nevertheless, one may still hear use of the term *natural experiment* applied to this type of serendipitous observation.

6.3 General concepts

The control group

Section 3 of Chapter 5 addressed the importance of using a referent group when judging the effects of an exposure. Referent groups in experimental studies are properly called **control groups**. Without the referent rate provided by the control group, it would often be impossible to determine the extent to which the rate in the treatment group reflected the effect of the treatment or the natural history of the disease.

In addition, when analyzing the results of a trial, the investigator is aware of the tendency of study participants to show improvements that are unrelated to the treatment being studied, at least temporarily. Several explanations have been advanced this phenomenon. Two such explanations are the placebo effect and the Hawthorne effect.

The **placebo effect** refers to perceived improvements following treatment with a pharmacologically inert substance (''placebo'') such as a sugar pill or saline injection. This effect has been ascribed to a positive belief in the treatment and the perception of being cared for.[a]

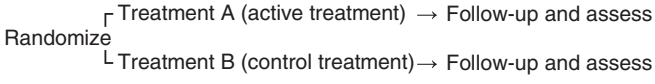
The **Hawthorne effect** refers to the tendency of subjects to alter their behavior in a way that is favorable to the results of the study. This effect was first described in a series of worker productivity studies conducted in the 1920s at the Hawthorne Works of the Western Electric Company in Chicago, IL, USA (Mayo, 1933). Continual improvements in worker performance were observed over the course of the study no matter the nature of the intervention. For example, worker output improved whether lighting was intensified or diminished. The Hawthorne and related effects have been attributed to the awareness of being observed and improved social conditions associated with observation. An **attention bias** analogous to the Hawthorne effect has been observed in subjects in health studies. A counter **John Henry effect** may occur when a control group getting no intervention compares themselves to the treatment group and responds by actively working harder to overcome the ''disadvantage'' of being in the control group (Sackett, 1979).

Because of factors such as the Hawthorne effect, it is important to compare the experience of the treatment group with that of a control group. It is also important to

[a] It has been suggested that approximately one-third of patients respond to a placebo (Beecher, 1955). Reconsideration of the placebo effect, however, suggests that most placebo effects can be explained by natural phenomena, such as spontaneous improvement, fluctuation of symptoms over time, regression to the mean, use of other unrecognized treatments, studying irrelevant response variables, answers of politeness and obsequiousness, conditioned responses, misjudgment, and misquotation (Kienle and Kiene, 1997).

care for and observe the treatment and control groups identically, to the extent that this is possible. **Blinding** (masking) the study participants and investigators about the treatment being received offers just such protection.

There are many different ways to incorporate control groups into an experiment. The simplest way is use a **parallel design** in which the experience of the treatment group and control group are compared concurrently:

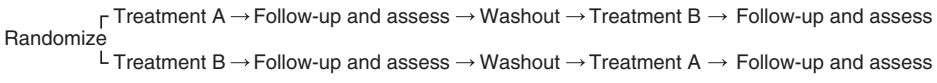


The importance of using a concurrent control group is demonstrated in this example.

Illustrative Example 6.3 Importance of concurrent control group (MRFIT)

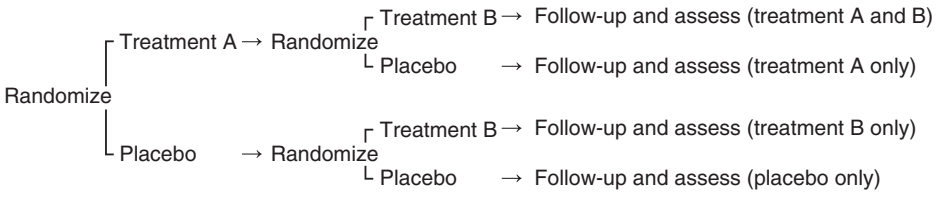
The Multiple Risk Factors Intervention Trial (MRFIT, 1982) was conducted in the early 1980s to test interventions intended to decrease the incidence of cardiovascular disease. Participants in the MRFIT trial were randomly assigned to either a special intervention group that received counseling or to a control group that received their usual sources of health care. After approximately 7 years of follow-up, the incidence of coronary disease mortality dropped precipitously in the special intervention group. However, it dropped equally in the control group. This is because the trial took place at a time when the entire country was learning about the benefits of reducing their cardiovascular disease risk profile by quitting smoking, decreasing dietary fat, and increasing exercise. Had the study included no control group, or had historical controls been used, it is likely that the intervention would unjustifiably have been declared a success resulting in the loss of millions of dollars on ineffective programs.

There are alternatives to simple parallel design for incorporating control groups into experiments. One such alternative is called a **cross-over design**. In a cross-over design, the treatment is first randomized. After a period of observation and measurement a ''washout'' period follows, during which the effects of the treatment subside. This is then followed by a cross-over to the alternative treatment by study subjects:



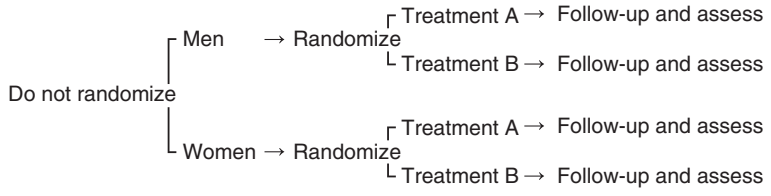
This creates a matched design where study subjects serve as their own ''control.''

More complex study designs are used to simultaneously assess the effects of two or more treatments. These are called **factorial designs**. For example, a factorial design may randomized multiple treatments sequentially, as follows:



Factorial designs may require sophisticated methods of analysis.

Finally, we consider **stratified** or **blocked designs** in which the treatment is randomized within separate strata or ''blocks.'' Here, for example, is a blocked design with blocks consisting of male and female strata:



Blocking on gender imposes a gender balance that might otherwise allude simple randomization, especially when the sample is small. For additional information on the design and analysis of the various types of trials, please refer to Pocock (1983).

Randomization and comparability

Randomization is the defining feature of modern trial design. By randomizing subjects to either a treatment or control group, extraneous factors that could otherwise confound the results of the study will tend to randomly distribute among the groups. When this process is effective, the effects of would-be confounders are neutralized.

The simplest way to randomize a treatment is to flip a coin as each subject is recruited into the study. If the coin turns heads up, the subject is assigned to the treatment group (for example). If it turns up tails, the subject is assigned to the control group. More elaborate randomization schemes are necessary when studying multiple groups and when more complex study designs are used. However, the principle remains the same: the luck-of-the-draw determines the treatment.

Illustrative Example 6.4 Randomized controlled trial (polio trial)

One of the most dramatic demonstrations of a field trial in modern history is the poliomyelitis trial of 1954. This field trial studied three-quarters of a million children at 211 different sites in the United States, Canada, and Finland (Francis *et al*., 1957). Individuals at these sites who requested participation in the study were registered and randomly given injections of either the Salk polio vaccine or a saline placebo.

Note that when the trial was initially proposed, it was suggested that everyone who agreed to participate be inoculated with the active vaccine while ''refusers'' serve as controls. Fortunately, cooler heads prevailed and a placebo-control group was included in the study. This was fortunate because it turned out that those who refused to participate were more likely to have lower incomes, have only one child, and be less likely to be exposed to the polio virus than those that were treated. The result was a lower than expected rate in the refusers, as demonstrated in these results:

- The vaccinated group had a polio rate of 28 per 100 000.
- The placebo group had a polio rate of 69 per 100 000.
- The ''refusers'' had a polio rate of 46 per 100 000.

Had the treatment group been compared with the ''refusers,'' the efficacy of the vaccine would have been underestimated by about half.

Checking group comparability

While randomization is of great benefit in creating group comparability, group differences can and do arise by chance, especially when the trial is small[b] or when the randomization procedure is flawed. In addition, the study intervention itself may inadvertently impart a group difference. For example, a study of a surgical intervention could introduce nonspecific effects of the operation. Therefore, even with the benefits of randomization, it is still important to compare the groups with respect to relevant cofactors that could affect the results of the experiment (Cornfield, 1954). The 1954 Polio Trial, for example, confirmed that the treatment and placebo group were comparable with respect to age, sex, race, the number of subjects who did not complete the protocol, the percentage with antibodies to natural polio virus before treatment, the distribution of titers to various other viruses, socioeconomic characteristics, and the extent of absenteeism from school (Francis, 1957). An additional example follows.

Illustrative Example 6.5 Baseline comparisons (WHI trial)

Table 6.1 compares selected characteristics of the treatment group and control group at the time that the study subjects were recruited to be in the Women's Health Initiative estrogen plus progestin trial. This table reveals no major differences between the groups at baseline, suggesting that the randomization procedure was effective. For those characteristics with a hint of a difference (e.g., prior treatment for coronary artery disease), the difference favored the estrogen plus progestin group, further strengthening the inference derived at the end of the study.

Recruitment and eligibility criteria

The **source population** for an experiment consists of those individuals who may potentially serve as study subjects. Identifying an appropriate source population requires setting up specific criteria for inclusion and exclusion. These criteria are called **admissibility (eligibility) criteria**. Well-defined admissibility criteria increase the suitability of study subjects while establishing a meaningfully homogenous source population by which to study the effects of the exposure.

Illustrative Example 6.6 Admissibility criteria (WHI trial)

The Women's Health Initiative referred to in earlier illustrative examples (e.g., Illustrative Example 6.1) consisted of healthy postmenopausal women between the ages of 50 and 79 years. Study subjects were recruited through direct mailings and media campaigns conducted at the 40 participating clinical centers. Recruitment took place between 1993 and 1998. Only predominantly healthy postmenopausal women with an intact uterus were eligible for participation. Potential participants were considered to be postmenopausal if they had experienced no vaginal bleeding for 6–12 months, depending on their age. Subjects were excluded from the study if they had a life threatening medical condition likely to present a competing risk during the study. In addition, subjects were excluded if they had characteristics that made adherence to the protocol either unsafe or uncertain (e.g., alcoholism,

[b] Randomization relies on the law of large numbers to balance the groups with respect to potential confounders. Small randomized experiments may therefore still suffer from group differences and confounding.

dementia). Of the 373 092 initially screened, 18 845 provided informed consent and had no prior hysterectomy. For women using postmenopausal hormones, a 3-month washout period was required before being randomized to one of the treatment arms. Ultimately, 16 608 women were randomized to either the treatment (n = 8506) or placebo (n = 8102) group.

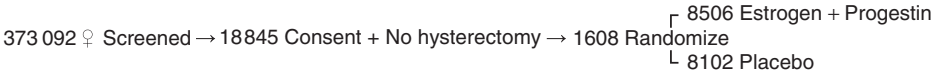


Table 6.1 Selected characteristics of participants in the WHI estrogen plus progesterone trial at entry into the study.

Characteristic	Estrogen + progestin (n = 8506)	Placebo (n = 8102)	P-Value
Age at screening, mean (SD), years	63.2 (7.1)	63.3 (7.1)	0.39
Age group (years)			
50–59	2839 (33.4)	2683 (33.1)	0.80
60–69	3853 (45.3)	3657 (45.1)	
70–79	1814 (21.3)	1762 (21.7)	
Race/ethnicity			
White	7140 (83.9)	6805 (84.0)	0.33
Black	549 (6.5)	575 (7.1)	
Hispanic	472 (5.5)	416 (5.1)	
American Indian	26 (0.3)	30 (0.4)	
Asian/Pacific Islander	194 (2.3)	169 (2.1)	
Unknown	125 (1.5)	107 (1.3)	
Hormone use			
Never	6280 (73.9)	6024 (74.4)	0.49
Past	1674 (19.7)	1588 (19.6)	
Current	548 (6.4)	487 (6.0)	
Body mass index, mean (SD)/kg m^{-2}	28.5 (5.8)	28.5 (5.9)	0.66
Systolic BP, mean (SD)/mm Hg	127.6 (17.6)	127.8 (17.5)	0.51
Diastolic BP, mean (SD)/mm Hg	75.6 (9.1)	75.8 (9.1)	0.31
Smoking history			
Never	4178 (49.6)	3999 (50.0)	0.85
Past	3362 (39.9)	3157 (39.5)	
Current	880 (10.5)	838 (10.5)	
Parity			
Never pregnant/no term pregnancy	856 (10.1)	832 (10.3)	0.67
≥1 term pregnancy	7609 (89.9)	7233 (89.7)	
Treated for diabetes	374 (4.4)	360 (4.4)	0.88
Treated for hypertension or BP ≥ 140/90 mm Hg	3039 (35.7)	2949 (36.4)	0.37
Elevated cholesterol levels	944 (12.5)	962 (12.9)	0.50
Statin use at baseline	590 (6.9)	548 (6.8)	0.66
Aspirin use (≥ 80 mg d^{-1}) at baseline	1623 (19.1)	1631 (20.1)	0.09
History of myocardial infarction	139 (1.6)	157 (1.9)	0.14
Prior surgical or percutaneous treatment for coronary artery disease	95 (1.1)	120 (1.5)	0.04

Source: WHI, 2002.

Follow-up and outcome ascertainment

The follow-up period of an experiment may be as brief as a few weeks for acute responses, or as long as years or decades for outcomes with long induction.

During the follow-up period, subjects are periodically assessed for the occurrence of relevant outcomes. Outcomes are ascertained during follow-up in a uniform manner using criteria established by **case definitions**.[c] Information is gathered through self-report, questionnaires, and direct examination. In the WHI Illustrative Example, for instance, study subjects were initially contacted by telephone 6 weeks after randomization to assess symptoms and reinforce adherence. Follow-up for clinical relevant study outcomes then occurred every 6 months. Yearly clinic visits were required.

Study outcomes are ascertained blindly whenever possible. **Blinding (masking)** is a study technique by which various parties are kept in the dark about the type of treatment each study subject is receiving. In the WHI Illustrative Example, study subjects did not know whether they were receiving the active treatment (estrogen plus progestin) or an identical looking placebo. Medication bottles were labeled with unique bottle number and bar code to allow for blinded dispensing. In addition, clinical evaluations were made without knowledge of the type of treatment each subject was receiving. This is a form of **double blinding** because two parties (study subjects themselves and the clinical evaluators) where masked about exposure status. **Triple blinding** may be applied by masking treatment assignments from additional parties such as the epidemiologists and statisticians responsible for interpreting the data.

Although blinding does not eliminate diagnostic errors, it encourages errors to balance among the study groups. Thus, misclassifications will tend to be ''nondifferential.'' Nondifferential misclassifications are preferable to differential misclassifications because they will tend to avoid spurious associations, as discussed in Section 9.3 (see especially Table 9.2).

Illustrative Example 6.7 Outcome ascertainment (WHI trial)

The WHI Illustrative Example included several study outcomes. However, the primary study outcome was myocardial infarction. The case definition for this outcome was established according to an algorithm adapted from standardized criteria based on hospitalization, cardiac pain, cardiac enzyme and troponin levels, and serial ECG readings (Ives *et al*., 1995). Cases were confirmed by specially trained physicians at each clinical center who were unaware of (''blinded to'') treatment assignments when they made their assessment.

Intention-to-treat analysis versus per-protocol analysis

Two different approaches exist for analyzing data from trials. These are intention-to-treat analysis and per-protocol analysis. **Intention-to-treat analysis (analyze-as-randomized, effectiveness analysis)** considers outcomes according to initially assigned treatments, irrespective of failures in compliance. In contrast, **per-protocol analysis (efficacy analysis)** considers outcomes only in participants who completed the treatment as intended by the study protocol.

[c] See Chapter 16 for additional information about case definitions.

Per-protocol analysis has the advantage of reflecting effects in those that received treatments as intended. Intention-to-treat analysis has two potential advantages that may not be initially evident. The first advantage of intention-to-treat analysis is that it more accurately reflects the way the treatment will perform under real conditions. The second advantage is that it simplifies the task of guarding against conscious and unconscious attempts to influence the results of the study.

Paul Meier provides a colorful example by addressing how investigator choice can adversely complicate the interpretation of trial results (Dallal, 1998). Suppose that during a period of treatment, a patient enrolled in a trial dies by falling off a boat after having been observed carrying a six-pack of beer on board the ship. Meier argues that most researchers would set this event aside as unrelated to treatment. In contrast, intention-to-treat analysis would require the death be counted against the treatment. Now suppose the beer is eventually recovered and every can is unopened. Intention-to-treat analysis would have done the right thing by including the case, whereas a study that permitted investigator judgment in this area adds a potential source of systematic error. Intention-to-treat analysis denies the investigator such leeway while encouraging errors to be randomly distributed among the treatment groups.

Illustrative Example 6.8 Intention to treat versus per-protocol analysis (WHI trial)

The WHI estrogen plus progestin trial used intention-to-treat principles as its primary basis for analysis. For a given outcome, the time of the event was defined as the number of days from randomization to the first diagnosis of the disease regardless of whether treatment was discontinued. Because some study participants discontinued study medications during follow-up, per-protocol analyses were also performed to examine the sensitivity of the intention-to-treat results. Per-protocol analyses censored a subject's experience 6 months after becoming nonadherent. The WHI investigators found that per-protocol analysis did not substantially affect the results of the intention-to-treat analysis, corroborating the conclusions of the study.

6.4 Data analysis

Measures of effect

Data analysis for experimental studies can range from simple comparisons of incidences to the use of survival analysis techniques (Chapter 17) and sophisticated regression models (e.g., Cox proportional hazards regression). This section will restrict itself to comparisons of incidence rates and proportions.

Recall that the effect of an exposure (treatment) can be quantified in relative terms or in absolute terms. The **relative effect** is quantified with this rate or risk ratio:

$$RR = \frac{R_1}{R_0} \tag{6.1}$$

where R_1 represents the incidence rate or incidence proportion (risk) in the treatment group, and R_0 represents the incidence rate or incidence proportion (risk) in the control group.

The **absolute effect** of the exposure is quantified by the rate or risk difference:

$$RD = R_1 - R_0 \tag{6.2}$$

Illustrative Example 6.9 Measures of effect (WHI trial)

The data for the WHI Illustrative Example found an incidence proportion of $\frac{164}{8506}$ or 0.019 28 for fatal and nonfatal coronary events in the active treatment group. The incidence proportion in the placebo group was $\frac{122}{8102}$, or 0.015 06. Thus, the $RR = \frac{0.019\,28}{0.015\,06} = 1.28$, indicating a 28% greater risk of coronary events in the active treatment group compared with the placebo control group.[d] During the period of follow-up, the *RD* was 19.3 per 1000 – 15.1 per 1000 = 4.2 per 1000, indicating an excess of 4.2 cases per every 1000 individuals in the active treatment arm of the study.

For treatments that are effective in curing or preventing illness, the *fraction of cases cured or prevented* is:

$$\text{Efficacy} = 1 - RR \tag{6.3}$$

Illustrative Example 6.10 Measures of effect (vitamin A trial)

Recall the vitamin A trial in the prevention of childhood mortality in Sumatra initially presented in Illustrative Example 6.2 (page 143). The childhood mortality rate in villages with vitamin A supplementation was 4.9 per 1000. The rate in the control villages was 7.3 per 1000. Thus, the $RR = \frac{4.8 \text{ per } 1000}{7.3 \text{ per } 1000} = 0.66$ and efficacy = 1 – *RR* = 1 – 0.66 = 0.34. Therefore, the program was 34% effective in reducing childhood mortality.

Statistical inference[e]

Effective randomization of the treatment in a trial permits the attribution of differences observed at the end of the study to either the treatment itself or to the role of chance in assigning the treatment. The role of chance is addressed with the standard techniques of statistical inference—*p* values and confidence intervals—both of which are easily calculated with computer applications such as *OpenEpi.com* (Dean *et al.*, 2011) and WinPEPI (Abramson, 2011).[f]

Illustrative Example 6.11 Inferential statistics (OpenEpi.com)

The WHI data in Illustrative Example 6.1 demonstrated 164 coronary events in the treatment group consisting of 8506 study subjects. The placebo group included 122 coronary events among 8102 study subjects. Figure 6.2 exhibits these data input into the *OpenEpi.com* →Counts →''Two by Two Table'' application.

[d] The 2002 WHI report derived virtually identical results with a statistical model that was able to address the time-to-events while simultaneously considering potential confounders such as clinical center, age of participant, prior disease, and status in a separate low-fat diet trial.

[e] It is assumed that readers will have had some statistical background before addressing this topic. For students with no prior introduction of statistical inference, section two in Chapter 9 can be covered before proceeding.

[f] *OpenEpi.com* is an open source website for epidemiologic statistics. WinPEPI (Windows Program for Epidemiologist) is a powerful public domain program that can be downloaded for free from *www.brixtonhealth.com*. See the article by Abramson (2011, *www.epi-perspectives.com/content/8/1/1*) for an introduction to WinPEPI.

Figure 6.3 displays *p* values derived by *OpenEpi.com* for testing the null hypothesis of ''no association.'' The mid-P exact method (highlighted by OpenEpi) derives a two-tailed *p* value of 0.037, suggesting that association cannot be easily explained by chance.

Figure 6.4 shows various risk, relative risk, and risk difference estimates. The *RR* estimate of 1.28 has a 95% confidence interval of 1.02 to 1.62. The *RD* estimate of 0.42% has a 95% confidence interval of 0.03–0.82%.

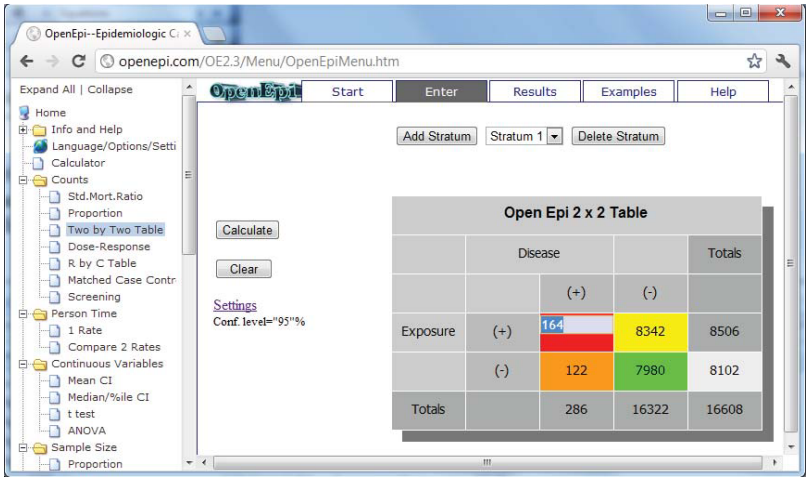


Figure 6.2 OpenEpi.com's input table for Illustrative Example 6.11.

Test	Value	p-value(1-tail)	p-value(2-tail)
Uncorrected chi square	4.372	0.01827	0.03655
Yates corrected chi square	4.126	0.02112	0.04224
Mantel-Haenszel chi square	4.371	0.01828	0.03655
Fisher exact		0.02095	0.04191
Mid-P exact		0.01828	0.03656

Figure 6.3 Screenshot of *p* values for Illustrative Example 6.11.

Point Estimates		Confidence Limits	
Type	Value	Lower, Upper	Type
Risk in Exposed	1.928%	1.656, 2.243	Taylor series
Risk in Unexposed	1.506%	1.262, 1.796	Taylor series
Overall Risk	1.722%	1.535, 1.932	Taylor series
Risk Ratio	1.28	1.015, 1.615[1]	Taylor series
Risk Difference	0.4222%	0.02766, 0.8168°	Taylor series

Figure 6.4 OpenEpi's measures of effect estimates for Illustrative Example 6.11.

Sample size requirements

Studying a sufficient number of subjects is essential to a trial's success. Therefore, sample size requirements for trials are determined before data are collected. Two approaches are used for calculating a sufficient sample size. When hypothesis tests of statistical significance are of primary importance, sample size requirements are based on detecting a specified difference with a given statistical power and alpha level. When the objective is to estimate an effect size, sample size requirements are based on the desired level of precision for a given level of confidence. In both instances, the sample size depends on expected rates of the outcome in the treatment and control groups and the ratio of the sizes of the two groups.

Discussions of sample size formulas are beyond the scope of this chapter and can be found in most biostatistics texts. However, calculations are readily achieved with WinPEPI (Abramson, 2011) and *OpenEpi.com*.

Illustrative Example 6.12 Sample size requirement (WinPEPI)

Figure 6.5 illustrates sample size calculations from WinPEPI →Compare2 →Sample size assuming an incidence proportion of 0.1 in the control group, incidence proportion of 0.2 in the treatment group, 80% power, an alpha level of 5%, and an allocation ratio of 1 control subject for each treated subject. The required sample size is 199 in each group to satisfy these conditions (with a regular chi-square test) and is 219 if a continuity correction factor is to be incorporated into the test statistic. The margin of error for the risk difference under these conditions will be ±0.075.

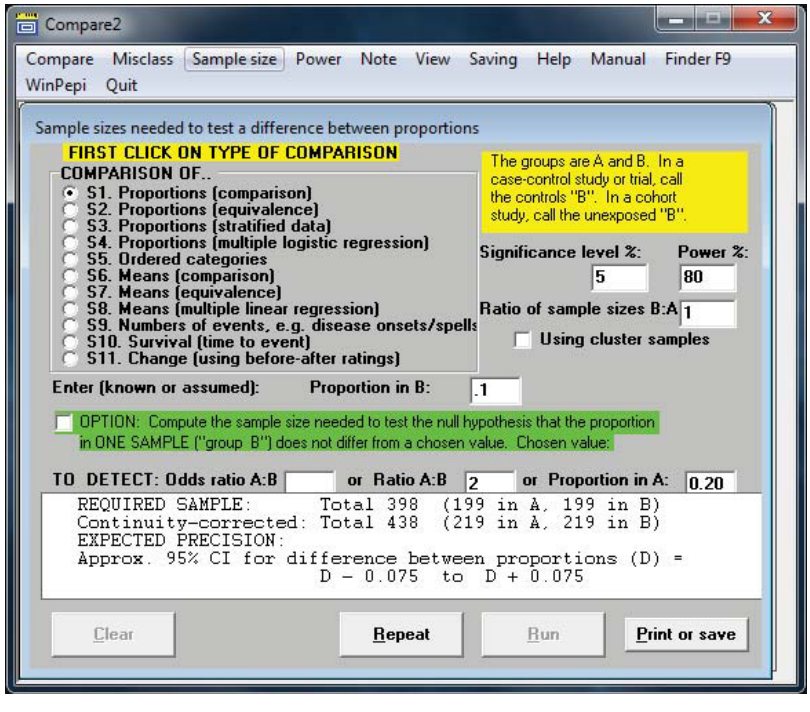


Figure 6.5 WinPEPI output for Illustrative Example 6.12; sample size requirements.

Exercises

6.1 Bicycle helmet campaign. You want to test whether a public awareness campaign about bicycle safety at elementary schools will increase bicycle helmets use among school-aged children. To test this intervention, you identify 12 elementary schools, half of which will be randomly assigned to participate in a school-wide bicycle helmet awareness program. The other 6 schools will serve as controls and will receive no special intervention. Research assistants will determine the percentage of bicyclists wearing helmets at standard locations in neighborhoods of each of the schools before and after the intervention.

(A) What is the unit of intervention in this study? (The ''unit of intervention'' refers to the level at which the intervention is randomized. This may differ from the ''unit of observation,'' which is the unit upon which the outcome is measured.)

(B) What is the unit of observation in this study?

(C) Even though the intervention was randomized in this study, there were only 6 treatment schools and 6 controls schools. Therefore, there is a good chance that treatment and control schools will differ with respect to important characteristics such as socioeconomic status. Can you think of a way to control for socioeconomic status through a randomization or study design approach?

6.2 Five cities. A population-based study examined changes in morbidity and mortality from 1979 through to 1992 in five northern California cities of moderate size (Fortmann and Varady, 2000). Two cities were delivered risk factor health education interventions through multiple educational methods targeted at all residents. Three of the cities served as controls. This was a nonrandomized trial. Fatal and myocardial infarction and strokes were identified from death certificates and hospital records in all five environments. Standard diagnostic criteria were used to classify events without knowledge of the city of origin. Population sizes (''denominators'') were derived from 1980 and 1990 US census figures. Overall, the combined-event rate declined about 3% per year in all five cities. The authors concluded that it is most likely that some influence affecting all cities, not the intervention, accounted for the observed change.

(A) What makes this study experimental and not merely observational?

(B) Explain why this is a community trial, *not* a field trial.

(C) Why were the myocardial infarctions and strokes identified without knowledge of the city of origin?

(D) Why was the inclusion of control cities essential for the proper interpretation of the results?

6.3 Fictitious HIV vaccine trial. Four hundred high-risk HIV-free volunteers were randomly assigned to either receive an experimental HIV vaccine or saline placebo. Subjects were randomized to form two equally sized groups of 200 each. Because of withdrawals, exclusions, and competing risks, 317 subjects completed the trial. After 5 years of observation, 4 HIV infections were observed in the treatment group (n_1 = 154 completing the study) and 11 HIV infections were observed in the control group (n_0 = 163 completing the trial).

(A) Is this study a clinical trial, field trial, community trial, or cohort study? Explain your response.
(B) List general and specific ethical concerns you might have in testing this vaccine.
(C) Calculate the per-protocol *RR* and efficacy statistic for the vaccine.
(D) Calculate the intention-to-treat *RR* and efficacy.
(E) Does the intention to treat analysis materially change your conclusion?

Review questions

R.6.1 The word trial comes from the French word *trier*. What does the *trier* mean in English?

R.6.2 What distinguishes experimental study designs from observational study designs?

R.6.3 What distinguishes clinical trials, field trials, and community trials?

R.6.4 Define each of these terms: randomized, controlled, double-blinded.

R.6.5 Explain why Lind's historical study of scurvy (Box 6.1) was an experiment even though it was not randomized.

R.6.6 Was Parés study of wound treatment (Section 6.2) truly an experiment?

R.6.7 Why do we observe the control group under the same conditions as the treatment group?

R.6.8 This term refers to ''improvements in behavior related to the fact that the study subjects know they are being observed.''

R.6.9 What is the placebo effect?

R.6.10 How does randomization reduce confounding?

R.6.11 What are admissibility criteria?

R.6.12 How does intention-to-treat analysis differ from per-protocol analysis?

R.6.13 What are the benefits of intention-to-treat analysis?

R.6.14 Provide a synonym for intention-to-treat analysis.

R.6.15 Why do we need evidence of group comparability even in randomized experiments?

R.6.16 List three reasons why randomization may not guarantee comparability.

R.6.17 Why is blinding useful even when outcome ascertainment is imperfect?

References

Abramson, J.H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations*, **8** (1), 1. Available: *http://www.epi-perspectives.com/content/8/1/1*. The latest version of WinPEPI can be downloaded from *www.brixtonhealth.com/*.

Armitage, P. (1983) Trials and errors: The emergence of clinical statistics. *Journal of Royal Statistical Society. Series A (General)*, **146**, 321–334.

Armitage, P., and Colton, T. (1998) Clinical trials, overview, in *Encyclopedia of Biostatistics* (eds P. Armitage and C.T. Colton) John Wiley & Sons, Inc., New York.

Beecher, H.K. (1955). The powerful placebo. *JAMA*, **159** (17), 1602–1606.

Cornfield, J. (1954) Statistical relationships and proof in medicine. *The American Statistician*, **8**, 11–13.

Dallal, G.E. (1998). Some aspects of study design in *The Little Handbook of Statistical Practice*, accessed 1 June 1 2001, from *http://www.tufts.edu/~gdallal/STUDY.HTM*.

Dean, A.G., Sullivan, K.M., and Soe, M.M. (2011) OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2.3.1, accessed July 2011, from *www.OpenEpi.com*.

Fortmann, S.P., and Varady, A.N. (2000) Effects of a community-wide health education program on cardiovascular disease morbidity and mortality: the Stanford Five-City Project. *American Journal of Epidemiology*, **152** (4), 316–323.

Francis, T. (1957). Symposium on controlled vaccine field trials: Poliomyelitis. *American Journal of Public Health*, **47**, 283–287.

Francis, T., Jr., Napier, J.A., Voight, R.B., Hemphill, F.M., Wenner, H.A., Korns, R.F., Boisen, M., Tolchinsky, E., and Diamond, E.L. (1957). *Evaluation of the 1954 Field Trial of Poliomyelitis Vaccine*, University of Michigan, Ann Arbor, MI.

Ives, D.G., Fitzpatrick, A.L., Bild, D.E., Psaty, B.M., Kuller, L.H., Crowley, P.M., *et al*. (1995) Surveillance and ascertainment of cardiovascular events. The Cardiovascular Health Study. [Comparative Study Research Support, U.S. Gov't, P.H.S.] *Annals of Epidemiology*, **5** (4), 278–285.

Kienle, G.S., and Kiene, H. (1997) The powerful placebo effect: fact or fiction? *Journal of Clinical Epidemiology*, **50** (12), 1311–1318.

Lilienfeld, A.M. (1982) Ceteris paribus: The evolution of the clinical trial. *Bulletin of History of Medicine*, **56**, 1–18.

Lind, J. (1753) A treatise on the scurvy. Retrieved 18 September 2012, from *http://www.bruzelius.info/Nautica/Medicine/Lind%281753%29.html*.

Mayo, E. (1933). *The Human Problems of an Industrial Civilization*, Viking Press, New York.

MRFIT (1982) Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial. Risk factor changes and mortality results. *JAMA*, **248**, 1465–1477.

Pocock, S.J. (1983) *Clinical Trials: A Practical Approach*, John Wiley & Sons, Ltd, Chichester.

Sackett, D.L. (1979) Bias in analytic research. *Journal of Chronic Diseases*, **32**, 51–63.

Sommer, A., Tarwotjo, I., Djunaedi, E., West, K.P., Jr., Loeden, A.A., Tilden, R., *et al*. (1986) Impact of vitamin A supplementation on childhood mortality. A randomised controlled community trial. *Lancet*, **1** (8491), 1169–1173.

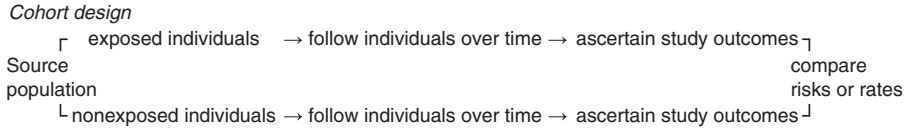
WHI (2002) Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*, **288** (3), 321–333.

CHAPTER 7
Observational Cohort Studies

7.1 Introduction
7.2 Historical perspective
7.3 Assembling and following a cohort
7.4 Prospective, retrospective, and ambidirectional cohorts
7.5 Addressing the potential for confounding
7.6 Data analysis
7.7 Historically important study: Wade Hampton Frost's birth cohorts
Exercises
Review questions
References

7.1 Introduction

The term **cohort** derives from the Latin word *cohors*, meaning ''an enclosure.''[a] This is an apt description of the cohort method because cohort studies follow the experiences of individuals enclosed in closed populations. In their simplest sense, cohort studies follow two groups of individuals. One group is characterized by an exposure and the other is ''nonexposed.'' Study outcomes in individuals are ascertained over time, tallied, and compared in the form of incidence rates or incidence proportions.



Cohort studies can be either experimental or observational (see Chapter 5 for the distinction). However, when used without specification, ''cohort study'' almost always refers to an observational cohort study. Chapter 6 addressed experimental cohort studies, while this chapter considers observational cohort studies.

[a] The Latin term *cohors* also applies to the tenth part of a Roman legion (a military unit of men).

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

Table 7.1 Six-year incidence of coronary heart disease according to initial serum cholesterol level, 40–59 year old Framingham Heart Study participants.

Serum cholesterol (mg/100 ml)	No. of incident cases	No. of individuals	Incidence proportion (%)
Men			
<210	16	454	3.52
210–244	29	455	6.37
≥245	51	424	12.03
Women			
>210	8	445	1.80
210–244	16	527	3.04
≥245	30	689	4.35

Source: Kannel *et al*. 1961.

Let us start our consideration of observational cohort studies with two examples.

Illustrative Example 7.1 Framingham Heart Study

Many groundbreaking findings about the causes of heart disease and stroke were identified and confirmed by the landmark Framingham Heart Study. The original Framingham cohort consisted of 5209 cardiovascular disease-free volunteers between the ages of 29 and 62 recruited from the moderately sized town of Framingham, Massachusetts, USA. The original Framingham cohort was recruited between 1948 and 1950. Individuals in the cohort have been examined every two years since the inception of the study. For the current illustration, cases of coronary heart disease (angina, myocardial infarction, and sudden death) were confirmed using criteria recommended by the New York Heart Association. Table 7.1 displays the data for 40–59 year old men and women for the first six years of follow-up. Note the progressive increases in coronary heart disease incidence with increasing serum cholesterol levels in both men and women.

Table 7.2 Breast cancer incidence by NSAID type and acetaminophen use.

Group	Duration of use at baseline (years)	*n*	Breast cancer cases	Person-yearsat risk	Incidence rate per 1000 person-years	*p* value for trend
Referent[a]	<1	54102	955	194884	4.90	N/A
Aspirin, ibuprofen, or prescription NSAID	1–4	9000	149	32127	4.64	0.01
	≥5	10162	148	36576	4.05	
Aspirin	1–4	5124	83	18231	4.55	0.03
	≥5	6759	99	24,398	4.06	
Ibuprofen	1–4	3469	51	12553	4.06	0.12
	≥5	2976	42	10653	3.94	
Prescription NSAIDs	1–4	1615	31	5552	5.58	0.21
	≥5	947	11	3388	3.25	
Acetaminophen	1–4	2450	44	8608	5.11	0.71
	≥5	4675	79	16698	4.73	

[a]The nonexposed referent category was composed of women in the cohort who reported less than one year of use of any NSAID or acetaminophen.
Source: Harris *et al*., 2003.

As a second example, let us consider an observational study from The Women's Health Initiative (WHI) project. The WHI project, introduced in Illustrative Example 5.2, included both experimental and observational cohorts. In Chapter 6, we discussed one of the experimental cohort studies. In this chapter, we consider one of its observational cohort studies.

Illustrative Example 7.2 NSAIDs and breast cancer

The recruitment period for the observational component of the WHI study began in September 1993 and ended in July 1998. Participants gave informed consent, were screened for eligibility, and were followed prospectively for up to 15 years (WHI, 1998). The current analysis explores the relation between non-narcotic pain medicine use and breast cancer (Harris *et al*., 2003). Information about the use of aspirin, ibuprofen, other nonsteroidal anti-inflammatory drugs (NSAIDs), and acetaminophen was collected from an interview-administered questionnaire.[b] For those individuals who reported use of an NSAID or acetaminophen at least two times in each of the two weeks preceding the interview, the type of compound, dose, and duration of use were recorded. The investigators checked pill bottle labels and prescription records to validate medication use.

Breast cancer cases were identified through health-care contacts and annual follow-up questionnaires. Follow-up time for each study subject was accrued from enrollment to the date of breast cancer diagnosis, death, or withdrawal from the study. Cases were confirmed by review of clinical, diagnostic, and pathology reports by physicians blinded to the exposure status of potential cases.

Table 7.2 summarizes findings from this study. An inverse dose–response relation between aspirin use, ibuprofen use, and NSAID use is evident. No such trend is found for acetaminophen use. This suggests that extended use of NSAIDs may reduce the risk of breast cancer.

Before addressing modern cohort studies further, let us gain an historical perspective into the method.

7.2 Historical perspective

The idea of observing people in their natural setting in order to gain insight into health determinants goes back a very long way. **Hippocrates** (circa 400 BCE to circa 370 BCE) urged us to consider the health of individuals in relation to ''the mode in which the inhabitants live, and what are their pursuits, whether they are fond of drinking and eating in excess, and given to indolence, or are fond of exercise and labor and not given to excess in eating and drinking'' (Hippocrates, 400 BCE). Here we recognize the seed of the observational cohort design, that is, comparing the long term health experiences of groups of individuals based on personal characteristics and exposures.

First century (CE) Roman authors noted differences in the patterns of diseases among various worker groups. Specific references to the diseases of slaves, sulfur workers, blacksmiths, and miners were made by the poets Pliny, Martial, and Luca (Rosen, 1993).

However, it was not until the 18th century when the focus on the health of worker cohorts met an important milestone with the publication of the treatise *De Morbis Artificum Diatriba* (''Diseases of Workers,'' 1713) by **Bernardino Ramazzini** (1633–1714).

[b] Aspirin and ibuprofen are non-prescription NSAIDs; acetaminophen (''Tylenol'') is not an NSAID.

De Morbis documented the effects of dozens of hazardous environmental exposures, such as specific chemicals, dusts, and abrasives. In addition, it documented the ill-effects of specific occupational practices and lifestyles. For example, the sedentary lifestyle and constant bent work posture of cobblers was cited by Ramazzini (1713) as a cause of ill-health.

In 1755, the surgeon **Percival Pott** (1713–1788) observed enormously elevated rates of scrotal cancer in chimney sweeps. He attributed this to the lodgment of soot in the rugae of the scrotum of the chimney sweeps, which is perhaps the first identification of an environment carcinogen.

Eighteenth century French physicians such as Louis and Pinel brought observational cohort studies into a clinical setting. In one study, **Pierre Charles Alexandre Louis** (1787–1872) observed superior cure rates in pneumonia patients who experienced delayed bloodletting compared with those who experienced early treatment. **Phillippe Pinel** (1745–1826) used the clinical histories of patients with mental illnesses to demonstrate superior cure rates at institutions that practiced humane methods of treatment compared with the standard inhumane practices of the time.

The Victorian physician and statistician **William Farr** (1807–1883), who is usually associated with vital statistics studies in open populations, also understood the importance of the longitudinal observation of individuals, declaring individual subjects in studies of health ''should be followed from the beginning to the end; every death or recovery should be recorded.'' Farr (1838) applied this concept in studying the prognosis of smallpox patients using sophisticated survival analysis techniques (Gerstman, 2003).

The 20th century brought with it the development of the modern cohort study. The British scientist **Janet Elizabeth Lane-Claypon** (1877–1967) reported on the longitudinal results of weight gain in infants fed either boiled cows' milk or human breast milk (1912). One year later, the German physician **Wilhelm Weinberg** (1862–1937) published the results of a large retrospective cohort study comparing the health experiences of 18 212 children whose fathers and mothers had previously died of tuberculosis with that of an ''nonexposed'' cohort of 7574 children of parents who died of non-tubercular causes (Morabia and Guthold, 2007).

In 1914, **Joseph Goldberger** (1847–1929) published cohort observations on pellagra, noting the absence of pellagra in nurses and ward attendants at hospitals in which 98 cases of pellagra had occurred in the patient population. Although the common belief at the time was that pellagra was infectious, this suggested to Goldberger that pellagra was not contagious.

It was not until 1935, however, that the first recorded use of the term *cohort* made its appearance in epidemiology when **Wade Hampton Frost** referred to rates of tuberculosis in generational cohorts (Doll, 2001). Frost's generational cohort studies are discussed in the last section of this chapter.

By the middle of the 20th century, the epidemiologic transition made it clear that large-scale long-term follow-up studies were needed in order to untangle the causes of chronic diseases in human populations. These modern cohort studies initially addressed cigarette-related diseases, cancers, and heart disease. One such study from this era—The Framingham Heart Study—was introduced as Illustrative Example 7.1. Another historically important cohort study from this era was the British Doctors Study.

Illustrative Example 7.3 The British Doctors Study

The British Doctors study was launched by Richard Doll and Bradford Hill in 1951 with a seven question survey form sent to approximately 59 600 medical doctors in the United Kingdom. Of the 59 600 mailings, roughly 50 000 were thought to have been received by living physicians. Of these, 34 440 physicians replied, representing a response rate of 69% of the men who were reached (Doll and Peto, 1976). A second questionnaire was sent out beginning in late 1957. By that time, 3122 of the cohort members had died, leaving 31 318 still alive. Of those remaining alive, 30 810 (98%) replied. By the time the third questionnaire was sent out in 1966, an additional 7301 had died. Of the remaining 27 139 living individuals, 26 163 (96%) replied to the survey. The fourth questionnaire, sent out in 1972, had a response rate of 98%. The nonresponse rates of 2, 4, and 2% are remarkably low, demonstrating the cooperative nature of the physicians that constituted the cohort.

The British Doctors cohort has now been followed for more than half a century. This study has identified or confirmed excess mortality in smokers due to dozens of neoplastic, vascular, and respiratory diseases (Doll *et al*., 2004). It also has also confirmed a negative association between smoking and Parkinson's disease (Doll *et al*., 1994).

7.3 Assembling and following a cohort

Before embarking on a cohort investigation, it is essential to obtain the **cooperation of the study population**. One of the reasons the landmark Framingham Heart Study was based in Framingham was because of its supportive population and health-care system (Dawber *et al*., 1963). The same can be said of The British Doctors Study. As another example, the Nurses' Health Study selected this population for long-term follow-up because nurses represented a large group of cooperative, health-conscious women who would be compliant and relatively easy to follow (Belanger *et al*., 1978).

However, even with a highly cooperative population, a certain percentage of individuals will refuse to participate. The aforementioned British Doctors, Framingham, and Nurses cohort studies had, for example, initial nonresponse rates that were all around 30%, (Doll and Peto, 1976; Dawber *et al*., 1963; Belanger *et al*., 1973).

In addition, subjects who enroll in a study may become lost to follow-up, refuse to continue in the study, or die from unrelated causes during the course of the study. These individuals are referred to as **withdrawals**. For example, the Framingham study began with 5209 subjects attending the initial set of examinations. Over the next ten years there were 950 withdrawals leaving the cohort with 4259 study subjects (Framingham Heart Study, 2011). This amounts to a ten-year withdraw rate of about 18%.

Nonresponses and withdrawals raise a potential for a ''selection bias'' in a cohort study. However, if nonresponse and withdrawal are independent of the exposures and diseases being studied, observed association will remain unbiased.

Once enrolled in a cohort study, each study subject is followed until:

- the subject withdraws from the study
- the study ends, or
- the study outcome is experienced.

The method of identifying the cases in a cohort study will depend greatly on the type of disease or health outcome being studied and the diagnostic and technologic resources that are available. Cases are ascertained based on criteria referred to as the **case definition**. The case definition is then uniformly applied to screen for and

confirm cases using relevant diagnostic and other information. See Chapter 16 for additional details regarding the construction of case definitions.

7.4 Prospective, retrospective, and ambidirectional cohorts

The cohort studies discussed to this point in the chapter were planned to observe events that had yet to occur. Cohort studies carried out in this manner are referred to as **prospective cohort studies**. Cohort studies can also be carried out using records of events that had occurred in the past. Studies of this second type are called **retrospective cohort studies** or **historical cohort studies**. Cohort studies that combine prospective data with retrospective data are called **ambidirectional cohort studies**. Figure 7.1 illustrates these temporal relationships.

Note that the study design feature that determines whether a cohort study is prospective or retrospective is the *proximity* in time of the data collection to the time events actually occurred. Prospective cohort studies used data that are proximal to data collection. Retrospective cohort studies use data from the past.

Retrospective data can be obtained from a variety of sources, including medical records, administrative data sources, vital records systems, surveillance systems, and employment records. In addition, we may interview study subjects or their proxies about prior events to obtain retrospective data. An example of a retrospective cohort study follows.

Illustrative Example 7.4 Retrospective cohort study of bladder cancer

Case and colleagues (1954) compiled a cohort of workmen in Great Britain based on historical records from 21 companies involved in the manufacture of aniline-based dyes. Among a total of 4622 men exposed to aniline dyes between 1921 to 1952, there were 127 mentions of bladder tumors on death certificates. Based on expected rates using national statistics for Britain as a whole, approximately four such death were expected in a group of this size and age distribution. Thus, the overall risk of dying of bladder cancer in the aniline-exposed cohort was approximately 30 times the expected rate.

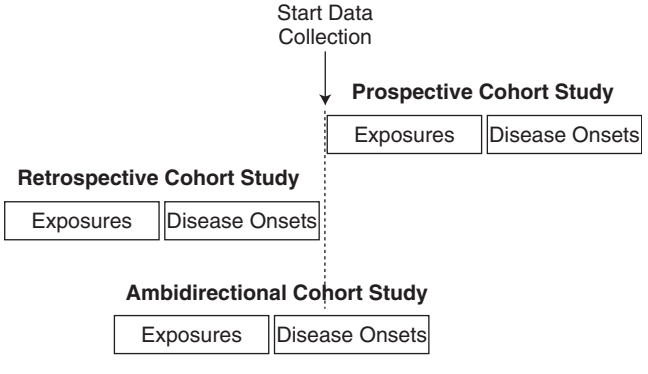


Figure 7.1 Proximity of data collection: prospective, retrospective, and ambidirectional cohort studies.

The above example illustrates an advantage of retrospective data: the investigator need not wait the many years required for disease to develop following exposure to a harmful substance to study outcomes with long induction. In addition, because the historical cohort studies use existing records, such studies can be completed relatively rapidly and economically.

7.5 Addressing the potential for confounding

The objective of a cohort study is to provide accurate information about the independent effects of study exposures on health outcomes. To accomplish this goal, like-to-like comparisons are necessary. Without like-to-like comparisons, differences in incidence found at the end of the study could be due to factors other than the study exposure. This phenomenon is known as **confounding**.[c]

Conceptually, the ideal nonexposed cohort would consist of the same individuals as the exposed cohort had these individuals been *non*exposed. Of course this is **counterfactual**, that is, not possible in fact. Nevertheless, this counterfactual ideal defines a way to think about suitable comparisons. To address this issue of comparability, cohort studies compare the distribution of risk factors and other attributes in exposure groups at the onset of the follow-up period. An example follows.

Illustrative Example 7.5 The Nurses' Health Study

The Nurses' Health Study is a prospective cohort study of married, female registered nurses born between 1 January 1921 and 31 December 1946. One of the many investigations that stemmed from this project addressed postmenopausal hormonal replacement therapy and cardiovascular disease. Table 7.3 compares the prevalence of cardiovascular risk factors among the exposure groups at enrollment. This table reveals that current hormone users were less likely to have diabetes, were leaner, and were more likely to engage in regular physical activity, to have had a surgical menopause, and to have used oral contraceptives in the past.

To encourage group comparability during the recruitment of study subjects, the investigator may restrict study subjects to those individuals with or without certain characteristics. This is achieved by imposing **admissibility (eligibility) criteria** when recruiting study subjects. For example, a study may admit only study subjects from a specific socioeconomic group. By imposing homogeneity with respect to this factor, confounding by socioeconomic status is averted.

Matching may also be used to encourage comparability. Two types of matching are considered: individual matching and frequency matching. Individual matching is achieved by matching study subjects on potential confounding during the recruitment process. For example, to control for age when studying the effects of smoking, we would match each 30-year old smoker with one or more similarly aged nonsmokers. Frequency matching, on the other hand, requires that the distribution of potential confounding factors be similar in exposed and nonexposed study groups. The intent

[c] Confounding is an observed difference in incidence in the exposed and nonexposed groups due to factors other than the study exposure. Confounding is discussed in greater detail in Chapter 9.

Table 7.3 Proportion of women in the Nurses' Health Study with coronary heart disease risk factors according to postmenopausal hormone use. The total sample size of the study was 48 470. Proportions have been adjusted for age.

	Estrogen use		
Risk factor	Current	Former	None
Parental heart attack before age 60	10.6	10.0	9.3
Hypertension	23.2	25.0	21.8
Diabetes mellitus	2.7	3.8	3.5
High serum cholesterol	9.9	11.2	7.6
Current smoker	11.2	14.7	14.5
BMI $\geq$29	9.8	13.3	15.0
Surgical menopause	50.3	39.3	9.3
Past use of oral contraceptives	34.0	27.6	23.9
Vigorous physical activity $\geq$1 time/week	48.2	43.1	42.4
Mean dietary intake saturated fats	27.6	26.2	26.7

Source: Stampfer *et al*., 1991.

of both types of matching is to create group comparability on the ''matched-on'' factor to avert the potential to confound.

When admissibility criteria or matching are not possible, the epidemiologists may then avail themselves of **statistical adjustments** to help mitigate (''control for'') the effects of potential confounders. Such adjustments may be based on stratified analysis or regression modeling. For example, the Nurses' Health Study cited in Illustrative Example 7.5 used a statistical modeling technique known as proportional-hazards regression to control for differences in age, cigarette smoking, hypertension, diabetes, high serum cholesterol, parental myocardial infarction, body mass index (BMI), past use of oral contraceptive, and time trends.

In summary, confounding can be averted in observational cohort studies through: (a) use of admissibility criteria during the recruitment of study subjects, (b) individual or group matching, or (c) statistical adjustment through stratification or regression modeling.

7.6 Data analysis

The fundamental epidemiologic measures of disease frequency and association were covered in Chapter 3. Let us illustrate several applications of such measurements in the current context. The primary analytic method of epidemiologic cohort studies is to calculate incidence proportions (risks) or incidence rates in cohort groups. These incidences can then be compared to determine associations between exposures and study outcomes.

The incidence proportion (risk) ratio and rate ratio (RR) measure the relative effect of an exposure by dividing the risk or rate in the exposed group (R_1) by the risk or rate in the nonexposed group (R_0):

$$RR = \frac{R_1}{R_0} \tag{7.1}$$

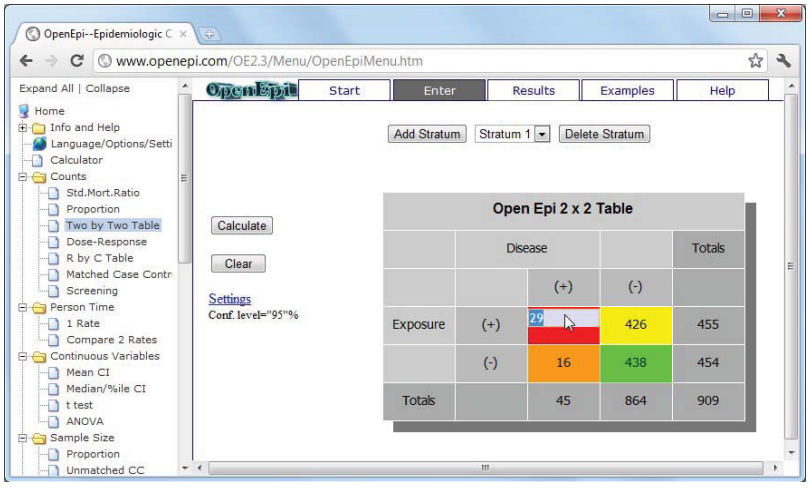


Figure 7.2 OpenEpi's input screen with data from Illustrative Examples 7.6.

This statistic quantifies the strength and direction of the association. An *RR* of 1 indicates no association between the exposure and disease. As the *RR*s gets further and further away from one in the positive (more than one) and negative (less than one) directions, the stronger and stronger the association (see Section 3.2).

Illustrative Example 7.6 Risk ratio (Framingham Heart Study)

Based on the data in Table 7.1, the 6-year incidence proportion (risk) of coronary heart disease in Framingham men between the ages of 40 and 59 with low serum cholesterol (<210 mg/dl) was 16 of 454, or 3.5%. The 6-year incidence in men with intermediate cholesterol (210–244 mg/dl) was 29 of 455, or 6.4%. Therefore, the $RR = \frac{R_1}{R_0} = \frac{29/455}{16/454} = \frac{0.0637}{0.0352} = 1.81$. This means that the men with an intermediate level of serum cholesterol had 81% greater risk (on a relative scale) than the men with low cholesterol.

Figure 7.2 shows the data in the 2-by-2-table program of *www.OpenEpi.com* (Dean *et al.*, 2011). Figure 7.3 shows the segment of OpenEpi's output that addresses risk estimates (i.e., estimates derived from incidence proportion data). The fourth line down shows a risk ratio of 1.809 with a 95% confidence interval of 0.9962–3.283. These results should be reported as either 1.8 (95% CI: 1.0–3.3) or 1.81 (95% CI: 1.00–3.28) to avoid the appearance of pseudo-precision.

Illustrative Example 7.6 exhibits a positive association between the exposure and disease using incidence proportions as the measure of disease frequency. Let us now consider an example with a negative association based on incidence rates.

Illustrative Example 7.7 Rate ratio (physical fitness and mortality)

A study of physical fitness and mortality found 25 deaths among 650 men who improved their fitness. The sum of the observation-times in this group of men was 4054 person-years. Therefore, their mortality rate was $\frac{25}{4054\,\text{person-years}} = 0.617\,\text{per}\,100\,\text{person-years}$.

In comparison, among 373 men who did not improve their fitness, there were 32 deaths during a total of 2937 person-years of observation. Therefore, the mortality rate in this group was $\frac{32}{2937 \text{ person-years}}$ = 1.09 per 100 persons-years (Blair *et al*., 1995).

Therefore, the effect of improved fitness expressed in relative terms is $\frac{0.617 \text{ per } \cancel{100 \text{ person-years}}}{1.09 \text{ per } \cancel{100 \text{ person-years}}}$ = 0.57. This rate ratio indicates a negative association between the exposure and study outcome, with improved fitness reducing mortality by 43% in relative terms.

The confidence interval for the rate ratio can be computed with *OpenEpi.com*'s ''Person-time Compare 2 rates'' program (Dean *et al*., 2009). Output for this example is shown in Figure 7.4. OpenEpi reports a rate ratio of 0.566 with a 95% CI of 0.3322–0.959. This should be reported either as 0.57 (95% CI: 0.33–0.96) or 0.6 (95% CI: 0.3–1.0) to avoid an appearance of pseudo-precision.

To quantify the excess risk or rate in absolute terms, calculate the incidence proportion (risk) difference or rate difference as:

$$RD = R_1 - R_0 \tag{7.2}$$

where R_1 is the risk or rate in the exposed group and R_0 is the risk or rate in the nonexposed group.

Illustrative Example 7.8 Risk difference (Framingham Heart Study)

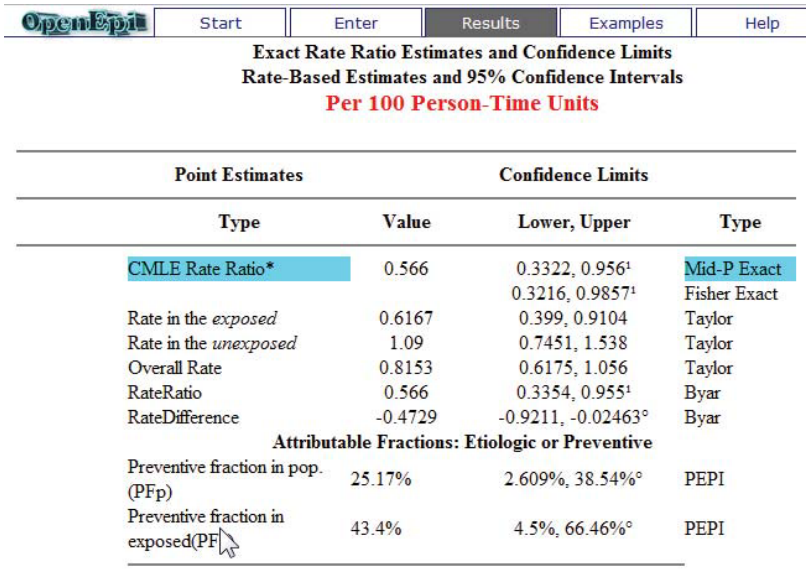
Recall the Framingham Heart Study data from prior illustrations, as shown in Table 7.1. The six-year incidence proportion of coronary heart disease in men with intermediate levels of cholesterol was 6.37% (R_1). The incidence proportion in the men with low cholesterol was 3.52% (R_0). Therefore, $RD = R_1 - R_0$ = 6.37% − 3.52% = 2.85%. This risk difference represents 2.85 additional cases per 100 people over six-years in association with intermediate cholesterol levels. The 95% confidence interval for this *RD* calculated with *OpenEpi.com* (see Figure 7.3) is 0.04–5.66%.

OpenEpi | Start | Enter | Results | Examples | Help

Risk-Based* Estimates and 95% Confidence Intervals
(Not valid for Case-Control studies)

Point Estimates		Confidence Limits	
Type	**Value**	**Lower, Upper**	**Type**
Risk in Exposed	6.374%	4.445, 9.033	Taylor series
Risk in Unexposed	3.524%	2.138, 5.69	Taylor series
Overall Risk	4.95%	3.708, 6.572	Taylor series
Risk Ratio	1.809	0.9962, 3.283[1]	Taylor series
Risk Difference	2.849%	0.0362, 5.663°	Taylor series
Etiologic fraction in pop. (EFp)	28.81%	1.42, 56.2	
Etiologic fraction in exposed(EFe)	44.71%	-0.383, 69.54	

Figure 7.3 OpenEpi's output screen with results for Illustrative Examples 7.6.


OpenEpi Start Enter Results Examples Help

Exact Rate Ratio Estimates and Confidence Limits
Rate-Based Estimates and 95% Confidence Intervals
Per 100 Person-Time Units

Point Estimates		Confidence Limits	
Type	**Value**	**Lower, Upper**	**Type**
CMLE Rate Ratio*	0.566	0.3322, 0.956[1]	Mid-P Exact
		0.3216, 0.9857[1]	Fisher Exact
Rate in the *exposed*	0.6167	0.399, 0.9104	Taylor
Rate in the *unexposed*	1.09	0.7451, 1.538	Taylor
Overall Rate	0.8153	0.6175, 1.056	Taylor
RateRatio	0.566	0.3354, 0.955[1]	Byar
RateDifference	-0.4729	-0.9211, -0.02463°	Byar
Attributable Fractions: Etiologic or Preventive			
Preventive fraction in pop. (PFp)	25.17%	2.609%, 38.54%°	PEPI
Preventive fraction in exposed(PF	43.4%	4.5%, 66.46%°	PEPI

Figure 7.4 OpenEpi.com's output for Illustrative Examples 7.7.

When applied to incidence rates, *RD* represents the rate difference.

Illustrative Example 7.9 Rate difference (physical fitness and mortality)

Recall the data from Illustrative Example 7.7 in which men who improved their physical fitness had a mortality rate of 0.617 per 100 persons-years (R_1), while men who remained unfit had a mortality rate of 1.09 per 100 person-years (R_0). Therefore, the $RD = R_1 - R_0$ = (0.62 per 100 persons-years) – (1.09 per 100 person years) = −0.47 per 100 person-years. Thus, the effect of improved fitness was to reduce mortality by almost $^1/_2$ case per 100 persons-years. The *OpenEpi.com* output in Figure 7.4 shows a 95% confidence interval for the rate difference of (−0.92 to −0.02) per 100 person-years.

For **ordinal exposures**, the ''least exposed'' group provides the ''non-exposed'' referent risk or rate, denoted R_0. Let R_k represent the risk or rate associated with the kth level of exposure. Thus, risk ratio or rate ratio associated with exposure level k is:

$$RR_k = \frac{R_k}{R_0} \tag{7.3}$$

The risk difference or rate difference associated with exposure level k is:

$$RD_k = R_k - R_0 \tag{7.4}$$

Illustrative Example 7.10 Multiple levels of exposure (Framingham)

The data for the men in Table 7.1 demonstrate the following relative risks:

- $RR_0 = \frac{R_0}{R_0} = \frac{3.52\%}{3.52\%} = 1$ (by definition)

- $RR_1 = \frac{R_1}{R_0} = \frac{6.37\%}{3.52\%} = 1.81$ (intermediate vs. low cholesterol)
- $RR_2 = \frac{R_2}{R_0} = \frac{12.03\%}{3.52\%} = 3.42$ (high vs. low)

The comparisons made in absolute terms are:

- $RD_0 = R_0 - R_0 = 3.52\% - 3.52\% = 0$ (by definition)
- $RD_1 = R_1 - R_0 = 6.37\% - 3.52\% = 2.85\%$ (intermediate vs. low)
- $RD_2 = R_2 - R_0 = 12.03\% - 3.52\% = 8.51\%$ (high vs. low)

Trends in risks and rates can be tested for statistical significance using option B of WinPEPI's Describe program (Abramson, 2011).[d]

Illustrative Example 7.11 Test for trend (NSAIDs and breast cancer)

Table 7.4 list breast cancer rates according to duration of NSAID use by women in the WHI observational cohort study (Harris *et al*., 2003). Let us use WinPEPI's Describe program (option B) to test whether the observed trend in these rates is statistically significant. Figure 7.5 exhibits WinPEPI's input screen. Output is displayed in Figure 7.6. The Mantel chi square statistic for trend is 4.68 with 1 degree of freedom, $P = 0.031$. This suggests that the observed trend is not easily explained by chance.

7.7 Historically important study: Wade Hampton Frost's birth cohorts

Wade Hampton Frost (Figure 7.7) was the first professor of epidemiology in the United States. He was also first to use the term ''cohort'' in an epidemiologic sense. By distinguishing open population rates from birth cohort rates, Frost was able to sort out a perplexing shift in the age distribution of tuberculosis in the population that had until then baffled epidemiologists.

Table 7.4 Breast cancer rates by duration of any NSAID use.

Duration of use	No. of cases	Person-years	Incidence rate per 1000 person-years	Rate ratio
<1 year (referent)	955	194884	4.90	$\frac{4.90}{4.90} = 1.00$
1–4 years	149	32127	4.64	$\frac{4.64}{4.90} = 0.95$
≥5 years	148	36576	4.05	$\frac{4.05}{4.90} = 0.83$

Source: Harris *et al*., 2003.

[d] The latest version of WinPEPI can be downloaded from *www.brixtonhealth.com*.

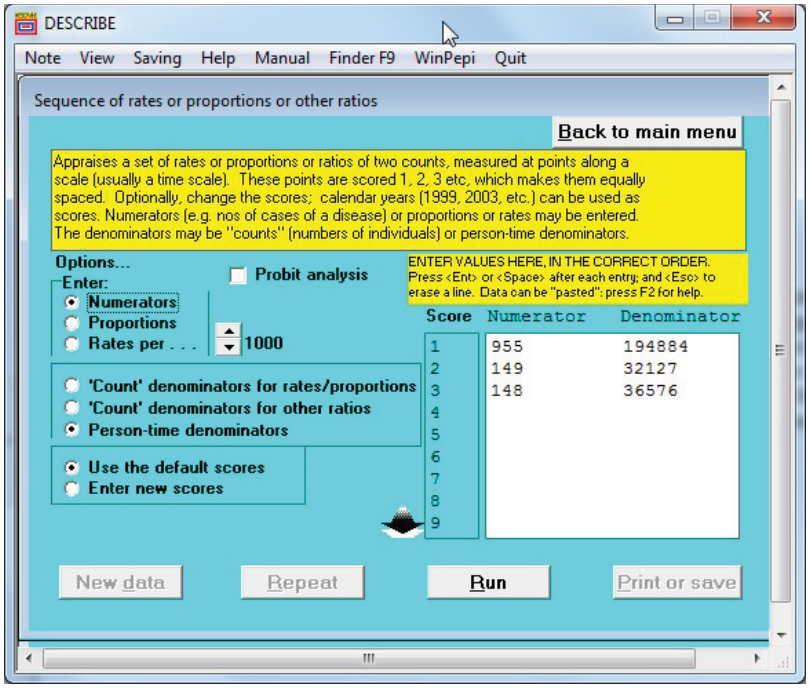


Figure 7.5 Input for Illustrative Example 7.11.

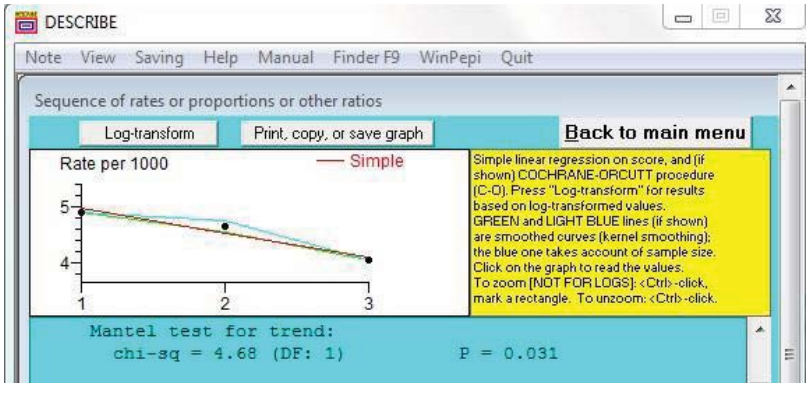


Figure 7.6 Output for Illustrative Example 7.11.

Table 7.5 exhibits data from Frost's posthumously published paper on tuberculosis mortality (Frost, 1939). Figure 7.8 plots open population ''cross-sectional'' rates from this table for the years 1880, 1910, and 1930. From this figure, the following trends are noted:

1 Decreasing rates over time, with age-specific rates highest in 1880 and lowest in 1930.

2 A peak in occurrence during early childhood (0–4 years) with a precipitous drop-off after age 5.

3 A peak occurring during adulthood (identified with a ¤ in Figure 7.8) that has shifted over time. In 1880, this adult peak was in 20–29-year olds. In 1910, the peak had shifted to age 40. In 1930, the peak has all but disappeared.

That tuberculosis rates were dropping (point 1) had been known for some time. The explanation for this was decreasing levels of the tuberculosis agent in the environment. In addition, the varying rates by age (point 2) could be explained by changes in host resistance with age or, as Frost put it, ''the balance established between the destructive forces of the invading tubercle bacillus, and the sum total of host resistance'' (1939, p. 92). The shifting peak in adulthood (point 3), however, was initially perplexing. Frost's insight into this shifting peak came in recognizing it as an artifact of open population rates. This was demonstrated by rearranging the rates to emulate the longitudinal experience of birth cohorts by reading the rates along diagonals of table. The tuberculosis experience of the 1880 birth cohort, for example, is shaded in Table 7.5.

When tuberculosis mortality rates are analyzed by birth cohorts, as in Figure 7.9, we can see that the relative frequency of tuberculosis mortality has been stable over time: the peak rate in adulthood remains in the 20s. There had been no change in age-related susceptibility over time. This finding had immediate relevance. There was some fear in the 1930s that postponement of infection to later ages was causing a more serious disease to occur, as is indeed the case with some infectious diseases such as measles and chickenpox. Some scientists and public health officials had even suggested that early exposure to the tuberculosis agent might afford immunologic benefits. Frost, on the other hand, believed that contact with the agent was to be avoided at all ages, supporting this view with his birth cohort data (Comstock, 2001).

Thus, the importance of distinguishing between short-term ''cross-sectional'' rates in open populations and long-term longitudinal rates in cohorts was established. This occurred at a critical time in history, as the epidemiologic transition was shifting the burden of disease in the population from acute infectious cause to chronic life-time cause, opening the door for a new, modern epidemiology.

Table 7.5 Tuberculosis mortality per 100 000 person-years by age, year, males, Massachusetts, 1880–1930. The experience of the 1880 birth cohort is shaded along the diagonal.

Age	1880	1890	1900	1910	1920	1930
0–4	760	578	309	209	108	41
5–9	43	49	31	21	24	11
10–19	126	115	90	36	49	21
20–29	444	361	288	207	149	71
30–39	378	368	296	253	164	115
40–49	364	336	253	253	175	118
50–59	366	325	267	252	171	127
60–69	475	346	304	246	172	95
70+	672	396	343	163	127	95

Source: Frost, 1939.

Figure 7.7 Wade Hampton Frost (1880–1938). Courtesy of Historical Collections & Services, Claude Moore Health Sciences Library, University of Virginia.

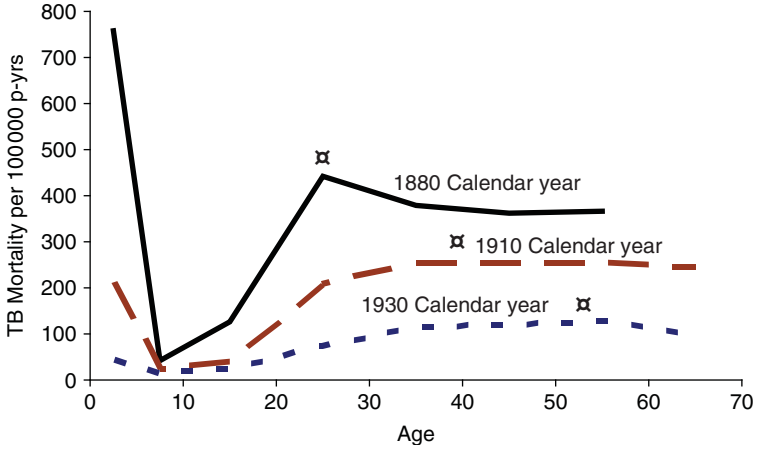


Figure 7.8 Cross-sectional tuberculosis mortality rates for men, calendar years 1880, 1910, and 1930; ¤ indicates peak rate in adults. (Source: Frost, 1939.)

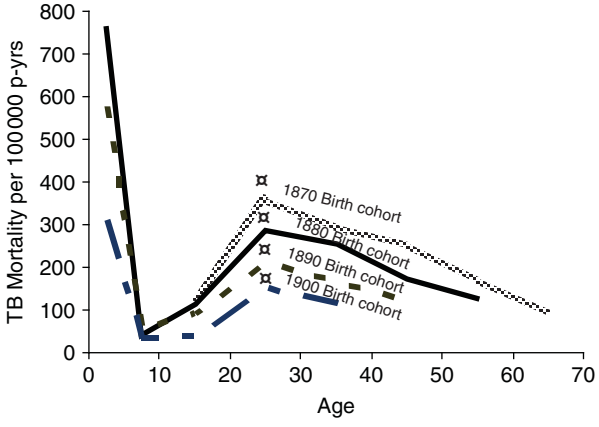


Figure 7.9 Tuberculosis mortality for the birth cohorts of 1870, 1880, 1890, and 1900 (Source: Frost, 1939).

Exercises

7.1 Agricultural related injuries. A cohort study of agriculture-related injuries among farm owners and farm workers from Alabama and Mississippi evaluated the experience of 685 Caucasian farm owners, 321 African–American farm owners, and 240 African–American farm workers (McGwin *et al*., 2000). Subjects were contacted biannually to ascertain the occurrence of agriculture-related injuries. The number of injuries and person-time of observation are:

Group	Agricultural-related injuries	Person-years of observation
Caucasian farm owners	67	2047
African–American farm owners	27	821
African-American farm workers	37	359

(A) Is this study prospective or retrospective in design?

(B) Calculate the rates of agricultural-related injury in the three groups. Then compare the rates as *RR*s using the Caucasian farm owners as the referent group.

(C) Based on these data, does race appear to be an independent risk factor for injury? Is being a worker compared to an owner an independent risk factor? Explain your responses.

7.2 NSAIDS and breast cancer. Illustrative Example 7.2 considered data from the Women's Health Initiative project, which followed 93 676 women between the ages of 50 and 79 from September 1993 to July 1998.

(A) Each study participant was required to sign a consent form before being enrolled in the study. Why was this important?

(B) All subjects were asked at admission if they had taken prescription or non-prescription pain medications. Checking pill bottle labels and prescription records validated medication use. Why was this important?

(C) Breast cancer cases were identified annually through follow-up questionnaires and health-care visits. Is this study a prospective, retrospective, or ambidirectional cohort study? Explain your response.

(D) Cases were confirmed by review or pathology reports, discharge summaries, operative reports, radiographic, and clinical findings. Coders were blinded to exposure status of potential cases. Why was blinding the coders important?

(E) Table 7.2 summarizes key results from the study. What is the *RR* associated with 1–4 years of aspirin use compared with the referent category?

(F) What is the *RR* associated with five or more years of aspirin use compared with the referent category?

(G) Summarize the results of this study with respect to aspirin use.

7.3 Oral contraceptive estrogen. A retrospective cohort study of venous thromboembolism in users of oral contraceptive products found that among users of formulations with 40 μg or less of ethinyl estradiol, 53 cases occurred during approximately 12.7×10^4 person-years of observation. Among users of formulations that contained 50 μg of ethinyl estradiol, there were 69 cases in 9.8×10^4 person-years of use (Gerstman *et al*., 1991).

(A) Calculate the rates of venous thromboembolism in each of the groups.

(B) Calculate the rate differences associated with each exposure level.

(C) How many additional cases of venous thromboembolism do you expect in women using the higher dose formulation compared with the low-dose formulations?

(D) This study used computerized diagnostic and treatment codes to identify cases. A separate validation study (Gerstman *et al*., 1990) based on medical record review was conducted to determine the reliability of this method of case ascertainment. During the validation study, medical record reviewers were blinded to the type of oral contraceptive used by suspected cases. Why was it important to blind the medical reviewers in this manner?

7.4 Anger prone personality and coronary heart disease.[e] A prospective cohort study by Williams and coworkers (2001) addressed whether people who angered easily were more likely to develop coronary heart disease than those who angered less easily. The Spielberger anger–temperament trait scale was used to classify each of the study subjects into low, moderate, and high anger-prone categories. Individuals were then followed for up to 72 months (median follow-up time 53 months) to ascertain the occurrence of coronary events.

---

[e] Based on a case study written by Vic Schoenbach, University of North Carolina.

(A) Identify the exposure variable in this study. Identify the study outcome.
(B) Explain why it was not possible to conduct an experimental study on this topic.
(C) A total of 14 348 study subjects were examined between 1990 and 1992. Of these, 1140 were determined to have a history of myocardial infarction, coronary bypass surgery, or electrocardiographic evidence of prior myocardial damage. These study subjects were excluded for further follow-up. Why?
(D) Approximately 93% of the study subjects returned for visit 2 of the study. How might the 7% loss to follow-up affect the study results?
(E) A comparison of the characteristics of study subjects at baseline suggested that high anger–temperament scores were related to gender (males greater than females), education (high anger trait less likely to have completed high school), cigarette smoking, alcohol use, plasma HDL cholesterol (lower in those with high anger trait), and wait-to-hip ratio (higher in those with high-anger trait). What are the implications of these differences in this study?
(F) Participants were followed from 31 December 1990 through to 31 December 1995. How many person-months would a study subject contribute to the cohort if they did not experience a coronary event and were not lost to follow-up?
(G) This table lists key information for normotensive and hypertensive study subjects. Calculate the incidence proportion of coronary events in low-anger normotensives (group 0), high-anger normotensives (group 1), low-anger hypertensive (group 2), and high-anger hypertensives (group 3). Discuss these findings.

Subjects	Normotensive		Hypertensive	
	Low anger (0)	High anger (1)	Low anger (2)	High anger (3)
No. of individuals at risk	8021	456	4231	282
No. of coronary events	167	23	213	13

7.5 Depression and Parkinson's disease. A study was conducted to determine whether depression is associated with Parkinson's disease. All patients visiting Northern California Veterans' Affairs hospitals between 1975 and 1990 were enrolled in the study and were evaluated for clinical depression (Shuurman *et al.*, 2002). Subjects were followed until 30 April 2000. Among the 1358 subjects diagnosed with clinical depression, 215 developed Parkinson's disease. Among the 57 570 non-depressed subjects, 1845 developed Parkinson's disease.
(A) Is this cohort study prospective or retrospective in design?
(B) Given only the information available in this exercise, what measure of disease frequency can be ascertained?
(C) Given only the information available in this exercise, what measures of association can be calculated?

(D) Calculate the risk ratio of Parkinson's disease associated with depression from the data provided. Return to the question addressed by this study and describe your results in this setting.

Review questions

R.7.1 Explain why rates derived from open populations are *not* longitudinal.

R.7.2 What does the Latin term *cohors* mean?

R.7.3 True or false? When an epidemiologist refers to a cohort study without specification they are almost always referring to an experimental cohort study.

R.7.4 Who wrote *De Morbis Artificum Diatriba* (''Diseases of Workers'')?

R.7.5 How was the first human environmental carcinogen discovered?

R.7.6 Joseph Goldberger (1847–1929) used cohort observations to hypothesize that pellagra was non-contagious. Explain Goldberger's reasoning.

R.7.7 Explain how Frost's generational cohort studies help modern epidemiology transition from a discipline primarily concerned with the study of acute diseases to one that is also concerned with the study of chronic diseases.

R.7.8 Where is Framingham and why is this town important epidemiologically?

R.7.9 Who are Doll and Hill?

R.7.10 How long has the British Doctors cohort study been followed?

R.7.11 Even with a highly cooperative population, a certain percentage of individuals invited to participate in a cohort study will fail to respond or refuse to participate. What was the initial nonresponse rate in The Framingham Heart Study?

R.7.12 The nonexposed group in a cohort study may be referred to as the referent group, while the exposed group may be referred to as the _____ group.

R.7.13 What determines whether a cohort study is prospective or retrospective.

R.7.14 Explain why *experimental* cohort studies are always prospective.

R.7.15 What is a synonym for ''retrospective cohort study''?

R.7.16 What is confounding?

R.7.17 Why do we compare the distribution of risk factors among groups in cohort studies at the onset of follow-up?

R.7.18 How does the use of admissibility criteria encourage group comparability?

R.7.19 Name three methods to mitigate confounding in cohort studies.

References

Abramson, J.H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations*, **8** (1), 1 *http://www.epi-perspectives.com/content/8/1/1*.

Belanger, C.F., Hennekens, C.H., Rosner, B., and Speizer, F.E. (1978) The nurses' health study. *The American Journal of Nursing*, **78** (6), 1039–1040.

Blair, S.N., Kohl, H.W. III, Barlow, C.E., Paffenbarger, R.S., Jr., Gibbons, L.W., and Macera, C.A. (1995) Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *JAMA*, **273**, 1093–1098.

Case, R.A.M., Hosker, M.E., McDonald, D.B., and Pearson, J.T. (1954) Tumors of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. *British Journal of Industrial Medicine*, **11**, 75–104.

Comstock, G.W. (2001) Cohort analysis: W.H. Frost's contributions to the epidemiology of tuberculosis and chronic disease. *Sozial- und Präventivmedizin (Social and Preventive Medicine)*, **46**, 7–12.

Dawber, T.R., Kannel, W.B., and Lyell, L.P. (1963) An approach to longitudinal studies in a community: the Framingham Study. *Annals of the New York Academy of Science*, **107**, 539–556.

Dean, A.G., Sullivan, K.M., and Soe, M.M. (2009, Version 2.3). OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2.3.1, accessed 29 December 2011 from *www.OpenEpi.com*.

Doll, R. (2001) Cohort studies: History of the method. I. Prospective cohort studies. *Sozial- und Präventivmedizin (Social and Preventive Medicine)*, **46**, 75–86.

Doll, R., and Peto, R. (1976) Mortality in relation to smoking: 20 years' observations on male British doctors. *British Medical Journal*, **2** (6051), 1525–1536.

Doll, R., Peto, R., Wheatley, K., Gray, R., and Sutherland, I. (1994) Mortality in relation to smoking: 40 years' observations on male British doctors. *British Medical Journal*, **309** (6959), 901–911.

Doll, R., Peto, R., Boreham, J., and Sutherland, I. (2004) Mortality in relation to smoking: 50 years' observations on male British doctors. [Research Support, Non-U.S. Gov't]. *British Medical Journal*, **328** (7455), 1519.

Farr, W. (1838) On prognosis. *British Medical Almanac, Supplement*, 199–216. Framingham Heart Study – Table of Exams. (2011). Retrieved 3 December 2011, from *http://www.framinghamheartstudy.org/share/index.html*.

Frost, W.H. (1939) The age selection of mortality from tuberculosis in successive decades. [Reprinted in the *American Journal of Epidemiology*, **14**1 (1), 4–9, (1995)]. *American Journal of Hygiene*, **30**, 90–96.

Gerstman, B.B., Freiman, J.P., and Hine, L.K. (1990) Use of subsequent anticoagulants to increase the predictive value of Medicaid deep venous thromboembolism diagnoses. *Epidemiology*, **1**(2), 122–127.

Gerstman, B.B., Piper, J.M., Tomita, D.K., Ferguson, W.J., Stadel, B.V., and Lundin, F.E. (1991) Oral contraceptive estrogen dose and the risk of deep venous thromboembolic disease. *American Journal of Epidemiology*, **133** (1), 32–37.

Gerstman, B.B. (2003) Comments regarding "On prognosis" by William Farr (1838), with reconstruction of his longitudinal analysis of smallpox recovery and death rates. *Soz Praventivmed (Social and Preventive Medicine)*, **48** (5), 285–289.

Goldberger, J. (1914) The Etiology of Pellagra: The Significance of Certain Epidemiological Observations with Respect Thereto. *Public Health Report*, **29** (26), 1683–1686.

Harris, R.E., Chlebowski, R.T., Jackson, R.D., Frid, D.J., Ascenseo, J.L., Anderson, G., *et al*. (2003) Breast cancer and nonsteroidal anti-inflammatory drugs: prospective results from the Women's Health Initiative. *Cancer Research*, **63** (18), 6096–6101.

Hippocrates. (circa 400 BCE) On airs, waters, and places. *The Internet Classics Archive*, accessed 21 June 2012, from *http://classics.mit.edu/Hippocrates/airwatpl.html*.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., and Stokes, J. III, (1961) Factors of risk in the development of coronary heart disease—six-year follow-up experience. *Annals of Internal Medicine*, **55**, 33–50.

Lane-Claypon, J.E. (1912) Report to the local government board upon the available data in regard to the value of boiled milk as a food for infants and young animals. Number 63. His Majesty's Stationary Office, London, United Kingdom.

Mantel, N. (1963) Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, **58**, 690–700.

McGwin, G., Jr., Enochs, R., and Roseman, J.M. (2000) Increased risk of agricultural injury among African-American farm workers from Alabama and Mississippi. *American Journal of Epidemiology*, **152** (7), 640–650.

Morabia, A., and Guthold, R. (2007) Wilhelm Weinberg's 1913 Large Retrospective Cohort Study: a rediscovery. *American Journal of Epidemiology*, **165** (7), 727–733.

Ramazzini, B. (1713; 2001 reprint) *De morbis artificum diatriba* [diseases of workers]. *American Journal of Public Health*, **91** (9), 1380–1382.

Rosen, G. (1993) *A History of Public Health (*revised edn*)*, Johns Hopkins University Press, Baltimore, MD.

Schuurman, A.G., van den Akker, M., Ensinck. K.T., Metsemakers, J.F., Knottnerus, J.A., Leentjens, A.F., and Buntinx, F. (2002) Increased risk of Parkinson's disease after depression: a retrospective cohort study. *Neurology*, **58** (10), 1501–1504.

Stampfer, M., Colditz, G., Willett, W., Manson, J., Rosner, B., Speizer, F., *et al*. (1991) Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the nurses' health study. *New England Journal of Medicine*, **325** (11), 756–762.

WHI (1998) Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Controlled Clinical Trials*, **19** (1), 61–109.

Williams, D. (1994) Epidemiology. Chernobyl, eight years on. (News). *Nature*, **371** (6498), 556.

Williams, J.E., Nieto, F.J., Sanford, C.P., and Tyroler, H.A. (2001) Effects of an angry temperament on coronary heart disease risk: The Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology*, **154** (3), 230–235.

CHAPTER 8
Case–Control Studies

8.1 Introduction
8.2 Identifying cases and controls
- Ascertainment of cases
- Selection of controls
- Number of controls per case
- Sample size considerations

8.3 Obtaining information on exposure
8.4 Data analysis
- Dichotomous exposure
- Multiple levels of exposure
- Matched pairs
- Matched tuples

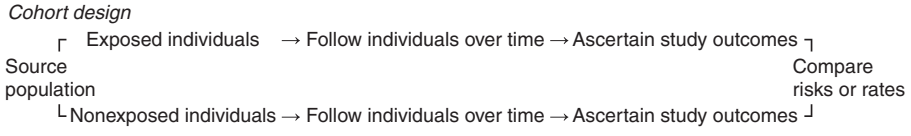
8.5 Statistical justifications of case–control odds ratio as relative risks
- Incidence density sampling
- Cumulative incidence sampling

Exercises
Review questions
References

8.1 Introduction

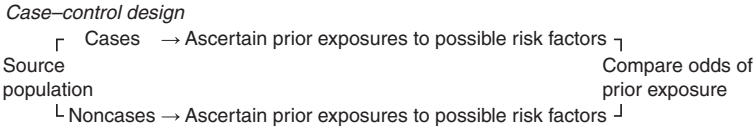
The two most common types of observational study designs in epidemiology are cohort studies and case–control studies. The objective of both these types of studies is to learn about causal relations between antecedent exposures and subsequent health outcomes. They differ, however, in the manner in which they select subjects for study.

Cohort studies start by identifying disease-free study subjects from a source population. Individuals are then classified as exposed or nonexposed and are followed (either prospectively or retrospectively) to determine incidents of relevant events.



Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

In contrast, **case–control studies** begin by identifying individuals who have already experienced the study outcome. These individuals comprise the case series. The study then selects a series of individuals *from the same source population* that gave rise to the cases but who have not (yet) experienced the study outcome. These individuals comprise the control series. Exposures to prior risk factors are then ascertained retrospectively in the cases and controls. The odds of prior exposures to risk factors are then compared in case series and control series.



Note that the distinction between cohort studies and case–control studies is based on the manner in which subjects are selected for study. Cohort studies start with disease-free study subjects and ''wait'' for the outcome to develop. Case–control studies start by selecting diseased and non-diseased subjects and ascertain prior exposures to risk factors. By selecting study subjects in this manner, case–control studies gain statistical efficiencies that could not otherwise be achieved when the disease is rare. In so doing, case–control studies forfeit the ability to estimate rates and risks directly (because the denominators needed to estimate risks and rates are absent). Case–control studies are, however, able to estimate the effect of an exposure in relative terms through a statistic known as the exposure **odds ratio**.

Table 8.1 exhibits the **notation** we use for cross-tabulated counts based on independent groups. In this notation, A represents ''case'' and B represents ''control.'' The subscript 1 represents ''exposure-positive'' and the subscript 0 represents ''exposure-negative.'' As examples, A_1 represents the number of exposed cases and B_0 represents the number of nonexposed controls. Using this notation, the odds of exposure in the case series is A_1/A_0, while the odds of exposure in the control series is B_1/B_0. The **exposure odds ratio** estimate[a] is thereby:

$$\widehat{OR} = \frac{A_1/A_0}{B_1/B_0} \tag{8.1}$$

This formula can be simplified to this algebraically equivalent form:

$$\widehat{OR} = \frac{A_1 B_0}{B_1 A_0} \tag{8.2}$$

Table 8.1 2-by-2 table notation for cross-tabulated counts based on independent samples.

	Cases	Controls
Exposure +	A_1	B_1
Exposure –	A_0	B_0

[a] The hat (ˆ) over the ''*OR*'' indicates that this formula provides an estimate of the underlying odds ratio in the population. The distinction between estimates and parameters will be made clear in Chapter 9.

Table 8.2 Data for Illustrative Example 8.1; alcohol consumption and esophageal cancer.

Alcohol (g/day)	Cases	Controls
≥80	96	109
<80	104	666
	200	775

Source: Tuyns *et al*. (1977).

Formula 8.2 is the **cross-product ratio** of the counts in the 2-by-2 cross-tabulation of counts: multiply the products of the counts in the cross-cells and form these products into a ratio.

It can be demonstrated that exposure odds ratios from case–control studies are direct estimates of the rate ratio in the underlying source population—see Section 8.5 for proof. Therefore, *OR*s are interpreted as if they were *RR*s. As examples, an *OR* of 1 indicates no association between the exposure and disease, whereas an *OR* of 2 indicates that the exposure doubles the risk.[b]

Illustrative Example 8.1 Esophageal cancer and alcohol consumption

A case–control study carried out in the Ille-et-Vilaine (Brittany) region of France identified 200 cases of esophageal cancer in men; 775 men without esophageal cancer were selected from electoral lists from the same region of France to serve as controls (Tuyns *et al*., 1977; Breslow and Day, 1980). Table 8.2 cross-tabulates counts from this study with alcohol consumption dichotomized at less than or more than or equal to 80 g/day. Note that 96 (48%) of the 200 cases are classified as exposed. In contrast, 109 (14%) of the 775 controls are classified as exposed. Thus, the $\widehat{OR} = \dfrac{A_1 B_0}{B_1 A_0} = \dfrac{96 \cdot 666}{109 \cdot 104} = 5.64$, indicating 5.64 times the risk of esophageal cancer in the high-alcohol consuming group relative to the low-alcohol consumers.

8.2 Identifying cases and controls

Ascertainment of cases

Before searching for cases, the study defines the diagnostic and epidemiologic criteria by which cases will be identified. These criteria are called the **case definition**. The case definition is then uniformly applied to screen for cases. At times it may be advantageous to establish several different case definitions to allow for separate examinations by the different criteria (see Chapter 16).

Cases ascertainment may be limited to incident cases or prevalent cases. **Incident cases** commenced during the study period. Prevalent cases may have onset at any time either before or during the study period. Incident cases are generally preferred

[b] See ''Relative Measures of Association'' in Section 2 of Chapter 3.

because the survival of prevalent cases may depend on factors that are separate from their cause. In addition, use of prevalent cases may make it difficult to differentiate between past events that are causally related to the disease and events that have occurred consequent to disease onset.

There are times, however, when it is possible only to study **prevalent cases**. For example, when studying birth defects, the causes occur *en utero*, before birth. Since many birth defects are associated with high fetal death rates and both spontaneous and induced abortions, birth defects detected after birth represent cases that have survived until delivery (i.e., prevalent cases). Note that use of prevalent cases will not bias the results of a case–control comparison if survival is independent of the cause you are studying and the timing of the exposure in relation to disease onset can be accurately ascertained.

Selection of controls

Valid selection of controls is crucial when conducting case–control studies. The valid selection is best understood in terms of their function, which is *to represent the exposure experience of the population that gave rise to the cases*. In this regard, it is important to clearly define the underlying **source population (study base)** that gave rise to the cases. As disease events arise in the study base, they are identified as cases. For each case, one or more controls are selected from the same study base.

The study base for a case–control study can be an opened or closed population (see Chapter 3). Cases for the study in Illustrative Example 8.1 (esophageal cancer and alcohol consumption), for example, came from hospitals in the open population of the Ille-et-Vilaine region (France). The **catchment area** served by these hospitals comprised the source population or study base. Therefore, controls were selected from this catchment area.

Case–control studies that use closed populations (cohorts) as their source population are called **nested case–control studies** (''case–control studies nested in a cohort''). As an example, a nested case–control study was carried out in a retrospective cohort of 223 292 men employed by three electric utility companies in France and Canada from 1970 to 1989 (Theriault *et al*., 1994). During this time, 4151 incident cancer cases were identified. For each incident cancer case, between one and four controls were selected from the underlying cohort. Exposure to electromagnetic field radiation for cases and controls was measured by dosimetry. Based on these data, workers who had more than the median cumulative exposure to magnetic fields had an increased risk of acute nonlymphoid leukemia ($\widehat{OR}$ = 2.4), acute myeloid leukemia ($\widehat{OR}$ = 3.2), and brain cancer ($\widehat{OR}$ = 2.0). No elevation in risk was observed for any of the other 29 cancers that were studied. These data strengthen the belief in the hypothesis that occupational exposures to 60 Hz electromagnetic fields increase the risk of certain types of cancer.

When cases came from a restricted source—say for example from a particular clinic—then the controls should be drawn from the same restricted source population. Here is an example in which an HMO population served as the source population for cases and controls.

Table 8.3 Data for Illustrative Example 8.2. Prostate cancer and vasectomy.

	Cases	Controls
Vasectomy +	61	93
Vasectomy –	114	165
	175	258

Source: Zhu *et al*. (1996).

Illustrative Example 8.2 Prostate cancer and vasectomy

A case–control study evaluated the relationship between vasectomy and prostate cancer in the Group Health Cooperative Health Maintenance Organization of Puget Sound (Zhu *et al*., 1996). Cases were 175 histologically confirmed cases of prostate cancer treated by the Health Maintenance Organization. The control series consisted of 258 similarly-aged men selected at random from the Health Maintenance Organization's general membership roles. Data were collected from medical record reviews and via questionnaires. Table 8.3 presents data for prior vasectomy in cases and controls. The $\widehat{OR} = \frac{A_1 B_0}{B_1 A_0} = \frac{61 \cdot 165}{93 \cdot 114} = 0.95$. Since this odds ratio is close to unity, it is reasonably interpreted as ''no association'' between vasectomy and prostate cancer.

This example illustrates an important benefit of the case–control approach. The study was completed with a total sample size of 433 (175 cases and 258 controls). Because prostate cancer is rare, occurring on the order of 150 cases per 100 000 man-years, a very large cohort would be required to derive 175 incident cases. This demonstrates the statistical efficiency of case–control studies.

Number of controls per case

In conducting case–control studies, the investigator must decide on the number of controls to select per case. When large numbers of cases are available and the cost and difficulty of collecting information for cases and controls is equal, maximum statistical efficiency is gained by studying an equivalent number of cases and controls. This control-to-case sampling ratio of 1:1 will maximize the statistical efficiency of the study for a given total sample size.

However, when the number of cases in the source population is limited, or the collection of information from cases is relatively expensive, an increase in the precision of the study can be achieved by increasing the control-to-case ratio sampling to 4:1. Increasing the control-to-case sampling ratio above 4:1, however, produces negligible increases in statistical power and precision (Gail *et al*., 1976).

Sample size considerations

The sample size requirements of case–control studies depends on the following factors: (a) the alpha level of the study, (b) the desired confidence interval length expressed on a logarithmic scale (or desired statistical power of the significance test), (c) the control-to-cases sampling ratio, (d) the expected proportion of controls that are classified as exposed, and (e) the expected odds ratio one is trying to detect.

Calculation of sample size requirements is facilitated with computer programs such as *www.OpenEpi.com* (Dean *et al*., 2009) and WinPEPI (Abramson, 2011).

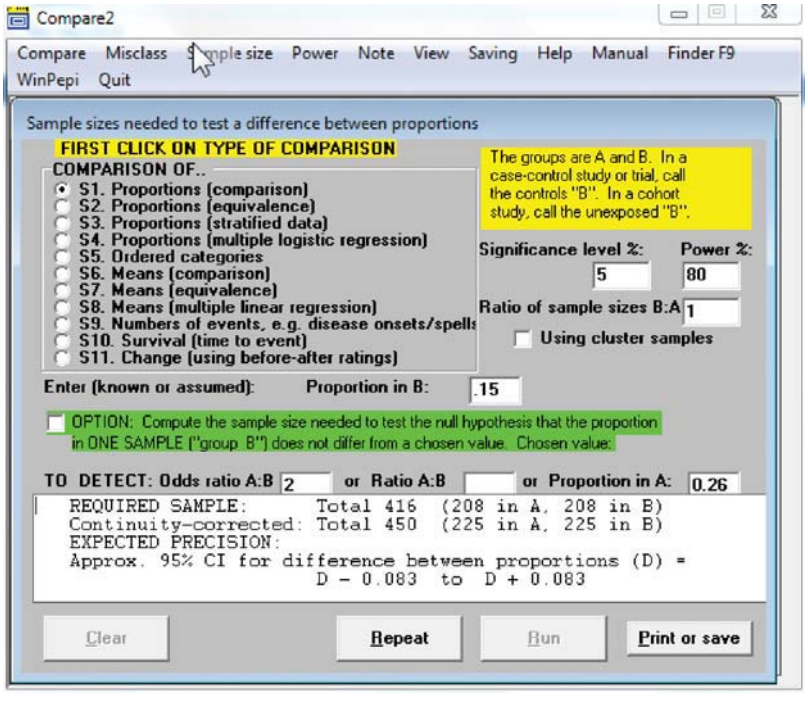


Figure 8.1 Sample size requirements for Illustrative Example 8.3 calculated by WinPEPI.

Illustrative Example 8.3 Sample size requirement, case–control study (WinPEPI)

Figure 8.1 is a screenshot from the program WinPEPI → Compare2 → Sample Size. This illustration indicates that a sample size of 208 cases and 208 controls is adequate to detect an odds ratio of 2 with 80% power at an alpha level of 5% when the prevalence of exposure to the risk factor in controls is 15%.

8.3 Obtaining information on exposure

Information about exposure to potential risks in cases-control studies can be derived by interviewing study subjects or their surrogates (e.g., family members), from the review of health care records, from vital statistics sources (e.g., death certificates), from employment records, from environmental records, and from biological specimens. Whatever the source, information should be obtained in a uniform and accurate way. For example, when deriving information by questionnaire, cases and controls should be questioned in identical manners to obtain information of comparable accuracy and completeness.

The induction time between exposure to a risk factor and its effects can be substantial for chronic diseases. In the meantime, cases may have altered their exposure habits since developing disease, making current exposure status irrelevant. For example, cases with chronic lung disease may have stopped smoking after the damage had

already been done. Therefore, it is important to focus on exposure histories from the relevant past.

Questionnaires and data collection forms should be thoughtfully worded and carefully designed. However, they need *not* be long and elaborate in order to be useful. For example, the questionnaire used to collect data for the landmark 1950 case–control study on smoking and lung cancer by Wynder and Graham (Figure 8.2) included only 15 items.

A key point in questionnaire design is to avoid misunderstanding of questions by study subjects and interviewers. We should neither take for granted that the interviewer nor interviewee will understand the question. The exposure being measured should be fully defined in each instance. Each question should be stated as precisely as possible. Items should be stated in a neutral way, taking care not to lead interviewees toward a particular response. If possible, the interviewer and interviewee should be kept unaware of the study hypotheses.

8.4 Data analysis

Dichotomous exposure

Illustrative Example 8.1 (Table 8.2) introduced the use of cross-tabulated data to derive the odds ratio associated with an exposures divided into two parts from a case–control study. Recall that this illustration derived an $\widehat{OR}$ of 5.6, indicating a strong positive association between esophageal cancer and alcohol consumption. Figure 8.3 is a screenshot of output for these data from the online application *www.OpenEpi.com* → Counts → two-by-two table. The odds ratio is reported as 5.627 with a 95% confidence interval of from 3.992 to 7.947. This should be reported as either 5.6 (95% CI: 4.0–7.9) or 5.63 (95% CI: 3.99–7.95) to avoid a spurious impression of precision.[c]

Multiple levels of exposure

Instead of dividing the exposure level in two, it is often helpful to classify the exposure according to multiple levels. For example, alcohol consumption can be divided into these four classes: 0–39, 40–79, 80–119, and 120+ g/day. Table 8.4 exhibits the data from the esophageal cancer and alcohol consumption illustrative example with the exposure classified thus.

When multiple exposure levels can be placed in rank order (as they are in Table 8.4), the least exposed group serves as the referent group for each comparison.[d] Then, using the notation listed in Table 8.5, the odds ratio associated with exposure level i is:

$$\widehat{OR}_i = \frac{A_i B_0}{B_i A_0} \tag{8.3}$$

[c] The term pseudoprecision has bee used to refer to the situation when an imprecise figure is given too many decimal places.

[d] When the exposure classes are categorical and cannot be placed in rank order, designation of the ''nonexposed'' referent group is arbitrary and does not materially affect the interpretation of results.

Name: _____ Age: _____

1. Have you ever had lung disease? Is so, state time, duration, and site of disease:

Pneumonia Asthma Tuberculosis Bronchiectasis

Influenza Lung Abscess Chest Injuries Other

2. Do you or did you ever smoke? Yes □ No □

3. At what age did you begin to smoke?

4. At what age did you stop smoking?

5. How much tobacco did you average per day during the past 20 years of your smoking?

Cigarettes _____ Cigars _____ Pipes _____

6. Do you inhale the smoke? Yes □ No □

7. Do you have a chronic cough which you attribute to your smoking, especially in the morning? If so, how long?

Yes □ No □ Duration _____

8. Do you smoke before or after breakfast? Before □ After □

9. Name the brand or brands and dates (if smoked for 5+ years).
First brand—from 19___ to 19___
Second brand—from 19___ to 19___

10. What kind of job have you held? Have you been exposed to dust or fumes while work there? (Use back of form for detailed descriptions of exposures.)

From	To	Position	Dust or fumes

11. have you ever been exposed to irritating dusts or fumes outside from you job? In particular have you used insecticide sprays excessively?

Yes □ No □ Type _____ Duration _____

12. How much alcohol do you average per day? State time & duration in years.
Whiskey _____ Beer _____ Wine _____

13. Where were you born and where have you lived most of your life (years)? Up to what grade did you attend school?
Birthplace _____ Home _____ Education level _____

14. State the cause of death of your parents, brothers and sisters.

15. *Site of Lesion* *Microscopic Diagnosis*
Papanicolaou Class *Etiologic Class*

Interviewer: _____

Figure 8.2 Facsimile of Wynder and Graham's (1950) data collection form. Notice its brevity.

Odds-based estimates and confidence limits

Point estimates		Confidence limits	
Type	**Value**	**Lower, upper**	**Type**
CMLE odds ratio*	5.627	3.992, 7.947[1]	Mid-p Exact
		3.937, 8.061[1]	Fisher Exact
Odds ratio	5.64	4.001, 7.951[1]	Taylor series
Etiologic fraction in pop. (EFp|OR)	39.49%	31.25, 47.73	
Etiologic fraction in exposed (EFe|OR)	82.27%	75, 87.42	

Figure 8.3 Output from *www.OpenEpi.com* for Illustrative Example 8.1 (Table 8.2), odds-based estimates.

Table 8.4 Esophageal cancer and alcohol consumption, case–control data.

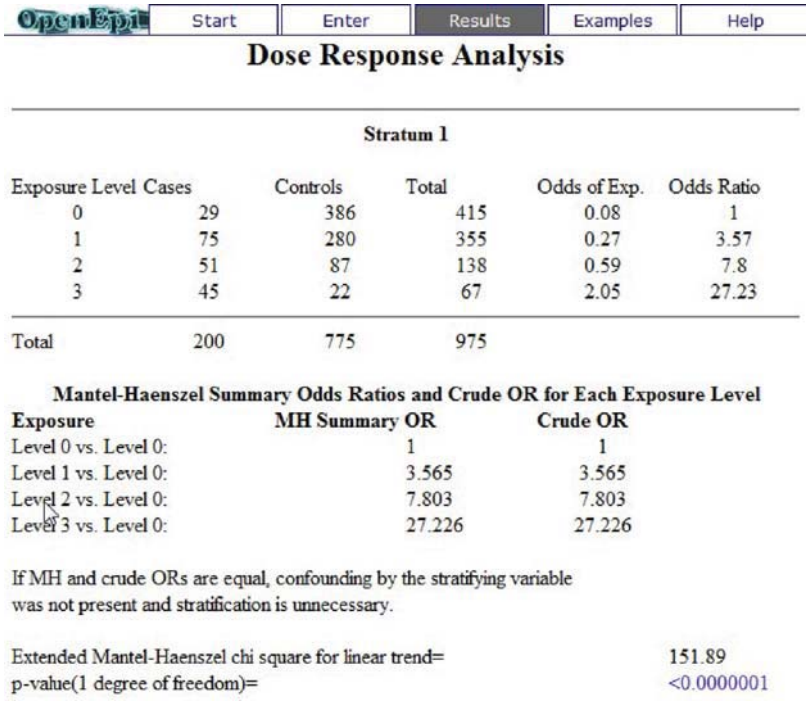
Alcohol (g/day)	Cases	Controls
0–39	29	386
40–79	75	280
80–119	51	87
120+	45	22
	200	775

Source: Tuyns *et al*. (1977).

Table 8.5 Notation for case–control studies with multiple levels of exposure[a].

Exposure level	Cases	Controls
0	A_0	B_0
1	A_1	B_1
i	A_i	B_i
⋮	⋮	⋮
k	A_k	B_k

[a]Exposure levels {i: 0, 1, . . . , k} are graded from low to high when possible. Level 0 represents the least exposed group.



OpenEpi | Start | Enter | Results | Examples | Help

Dose Response Analysis

Stratum 1

Exposure Level	Cases	Controls	Total	Odds of Exp.	Odds Ratio
0	29	386	415	0.08	1
1	75	280	355	0.27	3.57
2	51	87	138	0.59	7.8
3	45	22	67	2.05	27.23
Total	200	775	975		

Mantel-Haenszel Summary Odds Ratios and Crude OR for Each Exposure Level

Exposure	MH Summary OR	Crude OR
Level 0 vs. Level 0:	1	1
Level 1 vs. Level 0:	3.565	3.565
Level 2 vs. Level 0:	7.803	7.803
Level 3 vs. Level 0:	27.226	27.226

If MH and crude ORs are equal, confounding by the stratifying variable was not present and stratification is unnecessary.

Extended Mantel-Haenszel chi square for linear trend= 151.89
p-value(1 degree of freedom)= <0.0000001

Figure 8.4 Output from *www.OpenEpi.com* → Counts → Dose-response for Illustrative Example 8.4 (Table 8.4).

Illustrative Example 8.4 Esophageal cancer and 4-levels of alcohol exposure

For the data in Table 8.4:

- $\widehat{OR}_0 = \dfrac{A_1B_0}{B_0A_1} = \dfrac{29 \cdot 386}{386 \cdot 29} = 1$ (by definition)
- $\widehat{OR}_1 = \dfrac{A_1B_0}{B_1A_0} = \dfrac{75 \cdot 386}{280 \cdot 29} = 3.57$ (moderate vs. low alcohol consumption)
- $\widehat{OR}_2 = \dfrac{A_2B_0}{B_2A_0} = \dfrac{51 \cdot 386}{87 \cdot 29} = 7.80$ (high vs. low alcohol consumption)
- $\widehat{OR}_3 = \dfrac{A_3B_0}{B_3A_0} = \dfrac{45 \cdot 386}{22 \cdot 29} = 27.23$ (very high vs. low alcohol consumption)

The observed monotonically increasing odds ratios are suggestive of a biological gradient. This trend can be tested for statistical significance with the *www.OpenEpi.com* → Counts → Dose–response. Figure 8.4 exhibits the output from this program using the data in Table 8.4. The results reveal an extended Mantel (1980) chi-square statistic of 151.89 with 1 degree of freedom, P-value $< 0.000\,001$.

Matched pairs

Matching is used to help control for the potentially confounding effects of extraneous factors. For example, a case–control study that matches cases and controls on age will mitigate the potentially confounding effects of age in the study. However, matched data must be analyzed differently than unmatched (''independent'') samples.

Table 8.6 Notation for matched-pair case–control data.

	Case pair member	
Control pair member	Exposed	Nonexposed
Exposed	*t*	*u*
Nonexposed	*v*	*w*

Table 8.6 demonstrates the notation we will use for matched-pair case–control data. In this notation, table cell *t* represents the number of matched pairs in which both the case and control pair member are exposed to the risk factor, table cell *v* represents the number of matched pairs in which the case pair member is exposed and the control pair member is nonexposed, and so on. Notice that the counts in this table represent the number of pairs in the study, not the number of individual study subjects.

Also take note of the orientation of the notation table. In our notation, the exposure status of the *case* pair members is along the columns of this table and the exposure status of the *control* pair member is along the rows. Rotation of the table will not materially effect the interpretation of the data but will require adjustment of the odds ratio formula. Thus, the odds ratio for the matched-pair data is:

$$\widehat{OR} = \frac{v}{u} \tag{8.4}$$

Illustrative Example 8.5 Toxic shock syndrome and tampon use, matched pairs

In a matched-pair case–control study on toxic shock syndrome and tampon users, cases were asked to identify similarly-aged friends or acquaintances who lived in their region to serve as controls (Shands *et al*., 1980). Table 8.7 lists the tabulated results of the 44 matched pairs from this study. The exposure is ''continual tampon use,'' defined as using tampons every day and night throughout the menstrual period. The $\widehat{OR} = \frac{v}{u} = \frac{9}{1} = 9$ indicates that continuous tampon use was associated with nine times the risk of toxic shock syndrome compared with non-continuous use.

Figure 8.5 demonstrate the input screen with these data using ''WinPEPI → PairsEtc → A. 'Yes-no' (dichotomous) variable'' to calculate the odds ratio and associated statistics. Figure 8.6 exhibits the relevant part of the output for these data. The odds ratio point estimate and 95% confidence interval for the odds ratio by the exact mid-*P* method have been highlighted (Berry and Armitage, 1995). The results are reported as 9.0 (95% CI: 1.5–198.9).

Table 8.7 Continual tampon use and toxic shock syndrome, among matched pairs in which both the case and control used tampons.

	Case pair member	
Control pair member	Exposed	Nonexposed
Exposed	33	1
Nonexposed	9	1

Source: Shands *et al*., 1980, Table 4.

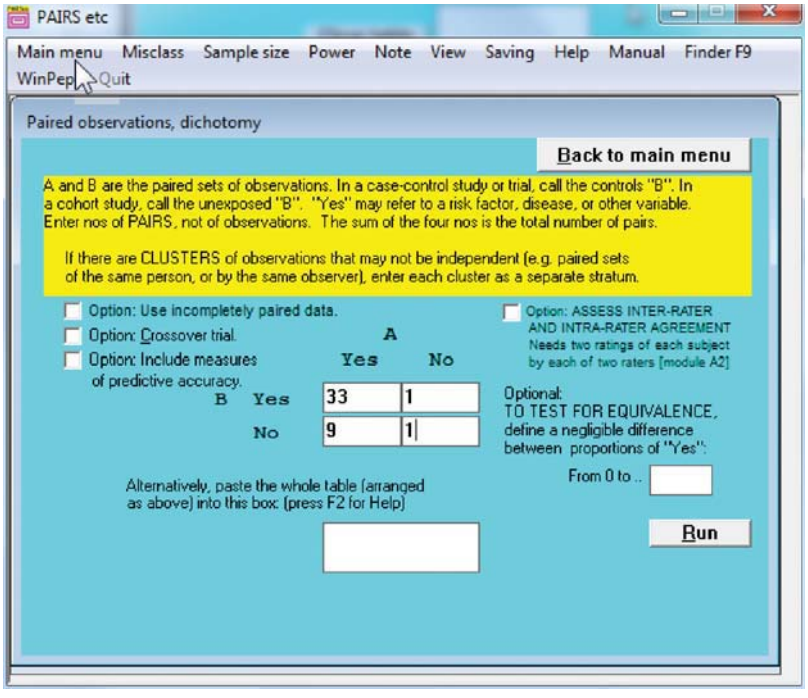


Figure 8.5 WinPepi input screen for Illustrative Example 8.5.

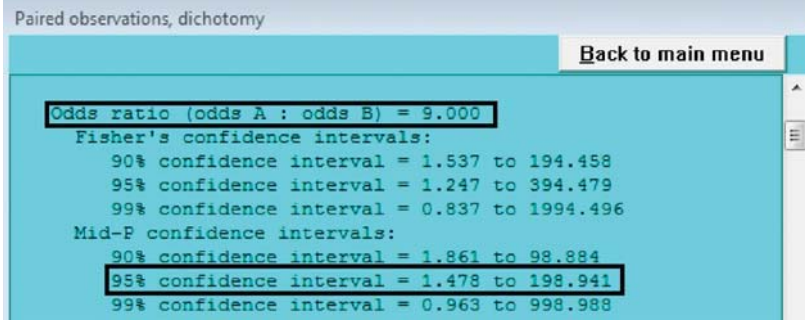


Figure 8.6 WinPepi output screen for Illustrative Example 8.5.

Matched tuples

Matched case–control studies may choose to match each case with more than one control. This increases the sample size and the precision of the odds ratio estimate with the odds ratio now calculated with this formula:

$$\widehat{OR} = \frac{\text{no.of nonexposed controls matched with an exposed case}}{\text{no. of exposed controls matched with a nonexposed case}} \tag{8.5}$$

To illustrate this approach, let us consider the following example.

Illustrative Example 8.6 Toxic shock syndrome and Rely brand tampons, matched four-tuples, three controls per case

Amid the publicity that surrounded the toxic shock epidemic of 1980 (addressed initially in Illustrative Example 8.5) there was added concern about a recently introduced highly absorbent brand of tampon called Rely. Therefore, in the fall 1980, the CDC launched a study to address whether one or more brands of tampons were more strongly associated with toxic shock syndrome than were other brands (Schlech *et al*., 1982). Much like in Illustrative Example 8.5, cases provide the names of female friends or acquaintances of approximately their same age who lived in their region to serve as controls. In this instance, each case identified three such controls, setting up sets of four, comprising three controls and one case.

One of the analyses from this study looked at Rely brand tampon use relative to other tampon brand use among tampon users. Table 8.8 lists the distribution of Rely brand use among 14 sets of four-tuples that met these criteria. Using Formula (8.5), the $\widehat{OR} = \frac{(1 + 10 + 12)\ \text{nonexposed controls matched to exposed cases}}{(0 + 2 + 1)\ \text{exposed controls matched to nonexposed cases}} = 7.67$.

Results can also be calculated with WinPEPI → PairsEtc → E. 'Yes-no' variable: compare subjects with 2 or more controls. Figure 8.7 exhibits the relevant part of the output for these data demonstrating an odds ratio or 7.66 and 95% confidence interval of 1.61–36.59.[e]

Table 8.8 Rely Brand tampon use and toxic shock syndrome among single brand tampon users, matched case–control data, three controls per case.

Rely brand use by case	No. of Rely users among controls	Number of sets	Total no. of discordances = no. of sets × discordances per set
Yes	3	1	1 × 0 = 0
Yes	2	1	1 × 1 = 1 (nonexposed control)
Yes	1	5	5 × 2 = 10 (nonexposed controls)
Yes	0	4	4 × 3 = 12 (nonexposed controls)
No	3	0	0 × 3 = 0 (exposed control)
No	2	1	1 × 1 = 1 (exposed control)
No	1	1	1 × 2 = 2 (exposed controls)
No	0	1	1 × 0 = 0

```
Mantel-Haenszel estimate      = 7.66
        S.E. of log. = 0.797
        90% conf. interval    = 2.06 to 28.46
        95% conf. interval    = 1.61 to 36.59
        99% conf. interval    = 0.98 to 59.80
```

Figure 8.7 WinPEPI output for Illustrative Example 8.6 (data from Table 8.8).

[e] Given the small number of cases, it may be more prudent to report the odds ratio and confidence limits with one decimal place accuracy, odds ratio = 7.7 (95% CI: 1.6–36.6).

8.5 Statistical justifications of case–control odds ratio as relative risks

Section 8.1 indicated that odds ratios from case–control studies are statistically equivalent to relative risks from cohort studies. Two justifications for this fact are presented. The first justification, referred to as incidence density sampling, demonstrates that odds ratios are stochastically equivalent to rate ratios in the underlying source population. The second justification is based on a different sampling method we will call cumulative incidence sampling.

Incidence density sampling

Figure 8.8 illustrates a principle of **incidence density sampling**. In this schematic, person No. 1 develops disease three years into the study. At that time, one of the four remaining disease-free individuals in the source population is selected at random to serve as a control. Thus, each disease-free individual has a 1-in-4 chance of selection at that time. A second case occurs at year 7, at which time an additional control is selected. Notice that case No. 2 might have also served as a control earlier in the study, although this would be unlikely if the disease has been rare.

Odds ratios derived by this method are stochastically equivalent to rate ratios in the underlying source population. To see this point, let A_1 represent the number of exposed cases in the source population, A_0 represent the number of nonexposed cases, T_1 represent the sum of exposed person-time, and T_0 represent the sum of nonexposed person-time. The rate ratio $= \frac{\text{rate}_1}{\text{rate}_0} = \frac{A_1/T_1}{A_0/T_0}$. This last expression here can be rewritten as $\frac{A_1/A_0}{T_1/T_0}$. The numerator of this last expression (A_1/A_0) represents the odds of exposures in the case series. The denominator (T_1/T_0) represents the odds of exposed to nonexposed person-time in the source population, which is estimated by the ratio of exposed controls to nonexposed controls (B_1/B_0). Therefore, the

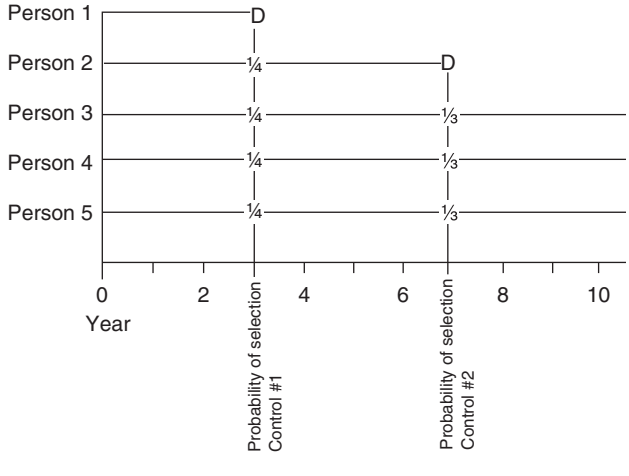


Figure 8.8 Incidence density sampling. *D* represents disease occurrence. Fractions represent probability of selection as a control at time *t*.

case–control odds ratio is a direct estimate of the rate ratio in the source population, no rare disease assumption required (Miettinen, 1976, 1985).

Cumulative incidence sampling

As with incidence density sampling, cumulative incidence sampling uses incident cases. However, selection of controls with cumulative incidence sampling is restricted to those remaining disease-free throughout the follow-up period.

Proof that the odds ratio from this case–control approach is statistically equivalent to a risk ratio requires two steps. The first step is a Bayesian demonstration of the equivalence of an incidence odds ratio and exposure odds ratio. In this proof, let Pr(D|E) denote the probability of disease given exposure, $\Pr(D|\bar{E})$ denote the probability of disease and nonexposure, Pr(E∩D) denote the joint probability of disease and exposure, and so on. If sampling is independent of exposure, then

$$\begin{aligned}
\text{Incidence odds ratio} &= \frac{\Pr(D|E) \,/\, \Pr\left(\bar{D}|E\right)}{\Pr\left(D|\bar{E}\right) \,/\, \Pr\left(\bar{D}|E\right)} = \frac{\Pr(D|E)}{\Pr\left(\bar{D}|E\right)} \times \frac{\Pr\left(\bar{D}|E\right)}{\Pr(D|E)} \\
&= \frac{\Pr(D \cap E) \,/\, \Pr(E)}{\Pr(D \cap E) \,/\, \Pr(E)} \times \frac{\Pr\left(\bar{D} \cap \bar{E}\right) \,/\, \Pr\left(\bar{E}\right)}{\Pr\left(D \cap \bar{E}\right) \,/\, \Pr\left(\bar{E}\right)} \\
&= \frac{\Pr(D \cap E)}{\Pr\left(\bar{D} \cap E\right)} \times \frac{\Pr\left(\bar{D} \cap \bar{E}\right)}{\Pr\left(D \cap \bar{E}\right)} \\
&= \frac{\Pr(D \cap E) \,/\, \Pr(D)}{\Pr\left(\bar{D} \cap E\right) \,/\, \Pr(D)} \times \frac{\Pr\left(\bar{D} \cap \bar{E}\right) \,/\, \Pr\left(\bar{D}\right)}{\Pr\left(D \cap \bar{E}\right) \,/\, \Pr\left(\bar{D}\right)} \\
&= \frac{\Pr(E|D)}{\Pr\left(\bar{E}|D\right)} \times \frac{\Pr\left(\bar{E}|\bar{D}\right)}{\Pr\left(E|\bar{D}\right)} \\
&= \text{Exposure odds ratio}
\end{aligned}$$

The second element of the proof requires a rare disease assumption. When the number of cases is small relative to the size of the population, the incidence odds ratio = $\frac{A_1/B_1}{A_0/B_0} \cong \frac{A_1/(A_1+B_1)}{A_0/(A_0+B_0)}$ = the incidence proportion (risk) ratio since A_1 and A_0 are miniscule relative to B_1 and B_0 when the disease is rare (Cornfield, 1951). Thus, the exposure odds ratio from a case–control study is equivalent to an incidence proportion (risk) ratio when the disease is rare.

Exercises

8.1 Influenza vaccination and primary cardiac arrest. It has long been recognized that influenza may precipitate death due to cardiovascular disease. To address this issue, a case–control study examined the relationship between influenza vaccination and primary cardiac arrest in King County, Washington, USA, between October 1988 and July 1994; 315 fatal cases of primary cardiac arrest were identified from paramedic reports. Among these cases, 79 had received an influenza vaccination during the prior 12 months. The researchers

used random digit dialing to contact 549 controls. Among 549 community controls, 176 were vaccinated against influenza (Siscovick *et al*., 2000).

(A) Explain why this is a case–control study and not, for example, a cohort study.

(B) Why do you think that the researchers used random digit dialing to identify controls?

(C) This study interviewed the spouses of cases and controls to ascertain exposure information. This was necessary because the case definition was for fatal primary cardiac arrest. Speculate on why the researchers chose to use spouses of controls to derive exposure information for the control series when the controls themselves were available for interview.

(D) Show the results of this study in two-by-two cross-tabulated form. Then calculate the odds ratio associated with vaccination and interpret your findings.

(E) Use *www.OpenEpi.com*, WinPEPI, or a comparable applet to calculate a 95% confidence interval for the odds ratio.

8.2 Historically important case–control study on smoking and lung cancer. The 27 May 1950 issue of the *Journal of the American Medical Association* included a case–control study on smoking and lung cancer by Wynder and Graham. Smoking was classified into five levels as follows:

5 Chain smokers (35 cigarettes or more per day for at least 20 years)
4 Excessive smokers (21–34 cigarettes per day for more than 20 years)
3 Heavy smokers (16–20 cigarettes per day for more than 20 years)
2 Moderately heavy smokers (10–15 cigarettes per day for more than 20 years)
1 Light smokers (1–9 cigarettes per day for more than 20 years)
0 Nonsmoker (less than 1 cigarette per day for more than 20 years)

Cross-tabulated results are as follows:

Smoking history	Cases	Controls
5	123	64
4	186	98
3	213	274
2	61	147
1	14	82
0	8	115
	605	780

Source: Wynder and Graham (1950).

(A) The authors noted ''we considered it particularly essential to learn how much a patient had smoked formerly, even though he might not smoke at all or smoke little at the time of the interview.'' Why was this important?

(B) Subjects were also queried about their occupation and whether they held jobs in which they were exposed to dust or fumes. Why was this important?

(C) Diagnoses of lung cancer were confirmed microscopically. Why was this important?

(D) Calculate the odds ratio associated with each level of smoking. Comment on your findings.

8.3 Thrombotic stroke in young women. A 1970s case–control study published in the *Journal of the American Medical Association* on stroke and oral contraceptives matched cases to controls on neighborhood, age, sex, and race. Matched-pair data from this study were as follows:

	Case pair member	
Control pair member	Exposed	Nonexposed
Exposed	2	5
Nonexposed	44	55

(A) Calculate the odds ratio of stroke associated with oral contraceptive use. Briefly interpret these results.

(B) How many cases used oral contraceptives?

(C) How many cases did not use oral contraceptives?

(D) How many controls used oral contraceptives?

(E) How many controls did not use oral contraceptives?

(F) Suppose the match was broken and investigators analyzed the data as if derived from independent cases and controls. Rearrange the information in the above table to show how it would appear in a standard 2-by-2 cross-tabulation. Then calculate the odds ratio for the data with the match broken. How does this improper analysis compare with the proper matched-pair analysis? (Note: The matched-pair table encompasses 106 pairs. Therefore, the 2-by-2 cross-tabulation will contain 212 independent counts.)

8.4 Esophageal cancer and alcohol consumption. This chapter included two illustrative examples from Ille-et-Vilaine case–control study of esophageal cancer. (Tuyns *et al.*, 1977; Breslow and Day, 1980). In these illustrative examples, we addressed alcohol consumption as a risk factor for this disease. The table below compares tobacco use in cases and controls.

Tobacco consumption	Cases	Controls
$\geq$ 30 g/day	31	51
$<$ 10 g/day	78	447

(A) Calculate the odds ratio for this comparison.

(B) Interpret the findings.

(C) Include a 95% confidence intervals.

(D) If cases tend to over-report tobacco consumption and controls under-reported it, how would this influence the results?

8.5 Pancreatic cancer and meat consumption. Among the 99 cases of pancreatic cancer in a case–control study, 53 indicated they had eaten fried or grilled meats at least weekly. Among 138 controls, 53 reported this behavior. Create a 2-by-2 table for these data. Then calculate the odds ratio and interpret the findings (Norell *et al*. 1986).

8.6 IUDs and infertility. A case–control study of infertility found prior use of a particular type of intrauterine device (IUD) in 89 of 283 in women with primary tubal infertility. In contrast, 640 out of 3833 fertile controls had used IUDs in the past (Cramer *et al*., 1985). Put the data in 2-by-2 table form, determine the odds ratio, and interpret the results.

8.7 Pedicure-associated furunculosis. A case–control study was initiated to examine the causes of pedicure-associated Mycobacterium furunculosis at nail salons in Santa Clara County, California, during the summer 2004. For this analysis, cases were those salons that were culture positive for acid-fast bacteria from environmental samples. Control salons were recruited randomly from a list of licensed nail salons that performed pedicures from the same county and were Mycobacterium furunculosis-free. Three different levels of cleaning were considered as potential risk factors: Factor A was lack of adequate cleaning at the end of the day; Factor B was lack of adequate cleaning after every client; and Factor C lack of extensive weekly cleaning. Cross-tabulations for each factor are shown in the tables below. Calculate the odds ratios associated with each factor to determine which are pertinent risk factors (Chung, 2010).

	Cases	Controls
Factor A +	8	1
Factor A−	10	10
Total	18	11

	Cases	Controls
Factor B +	12	3
Factor B −	5	8
Total	17	11

	Cases	Controls
Factor C +	5	1
Factor C −	11	9
Total	16	10

8.8 Case–control study of ZDV and HIV following needle sticks. Health-care workers are at risk of being exposed to HIV through needle sticks and other accidental punctures with sharp objects. A case–control study in the United Kingdom was conducted to determine the effects of administering zidovudine (ZDV) to health-care workers following needle stick injuries. The study included 31 health-care workers who became HIV positive (cases), of which 9 had received ZDV. A group of 679 HIV-negative health-care workers who also sustained needle sticks served as controls. Of these controls, 247 had received ZDV. Create a 2-by-2 table for these data and then calculate the odds ratio. Is there evidence that ZDV decreases the risk of HIV following needle stick injuries? If so, to what extent does ZDV prophylaxis cut down on risk HIV?

Review questions

R.8.1 What are the two most common types of observational studies in epidemiology?

R.8.2 What distinguishes case–control studies from cohort studies?

R.8.3 Why are case–control studies unable to determine the incidence of outcomes?

R.8.4 What measure of effect is available to case–control studies?

R.8.5 Fill in the blank: Odds ratios from case–control studies are stochastically equivalent _____ _____ in the underlying source population.

R.8.6 What do the subscripts 1 and 0 represent in the 2-by-2 table notation used in the text?

R.8.7 What do the symbols A and B represent in the 2-by-2 table notation?

R.8.8 This statistic is also called the cross-product ratio.

R.8.9 What does an odds ratio of 1 represent?

R.8.10 What does an odds ratio of 2 represent?

R.8.11 What does an odds ratio of 0.5 represent?

R.8.12 Identify sources of cases in case–control studies.

R.8.13 What is a ''case definition''?

R.8.14 What is the difference between an ''incident case'' and ''prevalent case''?

R.8.15 Identify sources of controls for case–control studies.

R.8.16 What is the primary function of the control series in a case–control study?

R.8.17 If cases were derived from a hospital, what is the best population-based source of controls?

R.8.18 What is a nested case–control study?

R.8.19 Why are case–control studies efficient for studying rare diseases?

R.8.20 Multiple choice: The odds ratio from a case–control study is stochastically equivalent to a(n): (a) incidence (b) prevalence (c) rate ratio (d) rate difference.

R.8.21 Multiple choice: When there is no association between the exposure and disease, the odds ratio is equal to (a) −1 (b) 0 (c) 1 (d) 100.

R.8.22 Fill in the blank: Maximum statistical efficiency for a given total sample size in a case-control study is achieved when we select _____ control for each case.

R.8.23 Fill in the blank: Increases in statistical power can be gained by selecting multiple controls per case in a case–control study. However, very little gain is achieved once we select more than __ controls per case.

References

Abramson, J.H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives and Innovation*, **8** (1), 1. *http://www.epi-perspectives.com/content/8/1/1*

Anscombe, F.J. (1956) On estimating binomial response relations. *Biometrika*, **43**, 461–464.

Belmont Report. (1979). *Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Retrieved 18 September 2012, from *http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html*.

Berry, G., and Armitage, P. (1995) Mid-P confidence intervals: a brief review. *The Statistician*, **44** (4), 417–423.

Breslow, N.E., and Day, N.E. (1980) *Statistical Methods in Cancer Research. Volume 1 – The Analysis of Case–control Studies*, International Agency for Research on Cancer, Lyon.

CDC (1980) Toxic shock syndrome—United States. *MMWR*, **29** (June, 27), 297–299.

CDC (1992) *Tampons and Toxic Shock Syndrome*. Association of Teachers of Preventive Medicine; Washington, DC.

CDC (1997) Toxic shock syndrome—United States. Editorial note—1997. *MMWR*, **46**, 492–495.

Cornfield, J. (1951) A method of estimating comparative rates from clinical data. *Journal of National Cancer Institute*, **11**, 1269–1275.

Cramer, D.W., Schiff, I., Schoenbaum, S.C., Gibson, M., Belisle, S., Albrecht, B., *et al*. (1985) Tubal infertility and the intrauterine device. *The New England Journal of Medicine*, **312** (15), 941–947.

Chung, J. (2010, April) Outbreak Investigation: Mycobacterial furunculosis at nail salons (Presentation, E-mails, and Data Files). San Jose.

Dean, A.G., Sullivan, K.M., and Soe, M.M. (2011, 2011/23/06) OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2.3.1. 2011/07/12, from *www.OpenEpi.com*.

Fleiss, J.L. (1981) *Statistical Methods for Rates and Proportions*, 2nd edn, John Wiley & Sons., Inc., New York, see page 64, formula 5.20.

Gail, M., Williams, R., Byar, D.P., and Brown, C. (1976) How many controls? *Journal of Chronic Disease*, **29** (11), 723–731.

Jewell, N.P. (1986) On the bias of commonly used measures of association for 2 × 2 tables. *Biometrics*, **42**, 351–358.

Levin, M.L., Goldstein, H., and Gerhardt, P.R. (1950) Cancer and tobacco smoking. *JAMA*, **143** (I), 336–338.

Mantel, N. (1980) Biased inferences in tests for average partial associations. *American Statistician*, **34**, 190–191.

Miettinen, O. (1976) Estimability and estimation in case-referent studies. *American Journal of Epidemiology*, **103**, 226–235.

Miettinen, O.S. (1985) The ''case–control'' study: Valid selection of subjects. *Journal of Chronic Diseases*, **38**, 543–548.

Norell, S.E., Ahlbom, A., Erwald, R., Jacobson, G., Lindberg-Navier, I., Olin, R., *et al*. (1986) Diet and pancreatic cancer: a case–control study. *American Journal of Epidemiology*, **124** (6), 894–902.

Payne, S. L. (1951) *The Art of Asking Questions*, Princeton University Press, Princeton, NJ.

Schlech, W.F., 3rd,, Shands, K.N., Reingold, A.L., Dan, B.B., Schmid, G.P., Hargrett, N.T., Hightower, A., Herwaldt, L.A., Neill, M.A., Band, J.D., and Bennett, J.V. (1982) Risk factors for development of toxic shock syndrome. Association with a tampon brand. *JAMA*, **248**, 835–839.

Shands, K.N., Schmid, G.P., Dan, B.B., Blum, D., Guidotti, R.J., Hargrett, N.T., Anderson, R.L., Hill, D.L., Broome, C.V, Band, J.D., and Fraser, D. W. (1980) Toxic-shock syndrome in menstruating women: Association with tampon use and *Staphylococcus aureus* and clinical features in 52 cases. *New England Journal of Medicine*, **303**, 1436–1442.

Siscovick, D.S., Raghunathan, T.E., Lin, D., Weinmann, S., Arbogast, P., Lemaitre, R. N., *et al*. (2000) Influenza vaccination and the risk of primary cardiac arrest. *American Journal of Epidemiology*, **152** (7), 674–677.

Theriault, G., Goldberg, M., Miller, A. B., Armstrong, B., Guenel, P., Deadman, J., Imbernon, E., To, T., Chevalier, A., Cyr, D., *et al*. (1994) Cancer risks associated with occupational exposure to magnetic fields among electric utility workers in Ontario and Quebec, Canada, and France: 1970–1989. *American Journal of Epidemiology*, **139**, 550–572.

Tuyns, A.J., Pequignot, G., and Jensen, O.M. (1977) [Esophageal cancer in Ille-et-Vilaine in relation to levels of alcohol and tobacco consumption. Risks are multiplying]. *Bulletin du Cancer*, **64**(1), 45–60.

Wynder, E.L., and Graham, E.A. (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *JAMA*, **143** (I) (4), 329–336.

Zhu, K., Stanford, J.L., Daling, J.R., McKnight, B., Stergachis, A., Brawer, M.K., *et al*. (1996) Vasectomy and prostate cancer: a case–control study in a health maintenance organization. *American Journal of Epidemiology*, **144** (8), 717–722.

Oral contraceptives and stroke in young women. Associated risk factors. (1975) *JAMA*, **231** (7), 718–722.

CHAPTER 9
Error in Epidemiologic Research

9.1 Introduction
- Random error and systematic error
- Parameters and estimates

9.2 Random error (imprecision)
- Probability
- Introduction to statistical inference
- Estimation (confidence intervals)
- Hypothesis testing (*p*-values)

9.3 Systematic error (bias)
- Selection bias
- Information bias
- Confounding

Exercises

Review questions

References

If you shut your door to all errors truth will be shut out.

Rabindranath Tagore (1916)

9.1 Introduction

Random error and systematic error

Effective use of epidemiologic information requires more than knowing the facts. It requires understanding the reasoning behind the methods. A good place to enter into this understanding is to view epidemiologic studies as **exercises in measurement** with the objective of measuring either disease frequency (e.g., an incidence rate) or association (e.g., a rate ratio) as accurately as possible. With this said, we acknowledge that all measurements are affected by varying degrees of random error and systematic error.

Random error and **systematic error** represent distinct problems in epidemiology. This distinction can be illustrated with a *metaphor*. Imagine a marksman shooting at a target in which the bull's eye represents the true value of the epidemiologic measure we want to learn about.

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

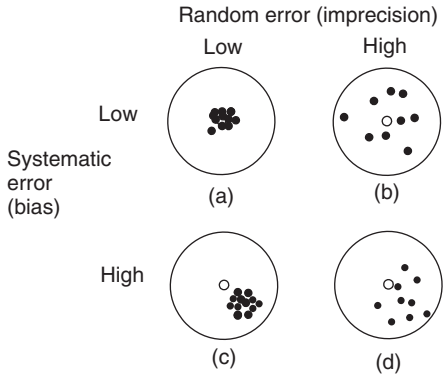


Figure 9.1 Target metaphor for random and systematic error.

- The skilled sharpshooter with a properly calibrated sighting device consistently delivers shots to the bull's eye center (Figure 9.1*a*). These shots are free from both random and systematic errors.
- If the sighting device of the gun is true, but the sharpshooter is forced to shoot from a randomly vibrating surface, the shots are scattered about the target (Figure 9.1*b*). These shots are free from systematic error but are affected by random error.
- If the sighting device of the gun is askew and the shooter is on stable, even ground, the shots will be consistently off-center (Figure 9.1*c*). These shots are free from random error but are affected by systematic error.
- If the instrument is improperly calibrated and the sharpshooter is forced to shoot from a vibrating surface, shots distribute themselves randomly *and* systematic error is present (Figure 9.1*d*).

With random error, each particular shot is unreliable but, *on the average*, the shots tend to center on the bull's eye. With systematic error, shots are systematically off-center in a particular direction.

Parameters and estimates

Instead of considering sharpshooters, let us now consider epidemiologists ''shooting'' for the correct value of an epidemiologic measure. When calculating an incidence, for example, the epidemiologist is seeking the error-free value of the incidence in the population being studied. When calculating a risk ratio, (s)he is seeking the absolutely correct value of the risk ratio that describes the exposure–disease relation of interest.

We will refer to the error-free value of the epidemiologic measure as the **parameter**. For instance, when the epidemiologic measure of interest is a risk ratio, the risk ratio parameter quantifies the true effect of the exposure on the occurrence of the disease in relative terms.

Although parameters are objective characteristics of the population being studied, they are impossible to observe (read: calculate) directly. Instead, we calculate an imperfect **estimate** of the parameter based on the data from a study. This imperfect estimate is prone to both random and systematic errors. Thus, the parameter is

analogous to the bull's eyes of the targets in Figure 9.1, while a given estimate is analogous to a single shot.

Accordingly, we require different **notation** to preserve the distinction between the parameter we are trying to estimate and the estimate itself. In general, estimates will now carry an overhead hat (ˆ) while parameters will remain hatless. For example, $\widehat{MA}$ will represent a measure of association estimate from a particular study. In contrast, MA will represent the (error-free) measure of association parameter.

We can now think of independent studies labeled 1 through k deriving independent measures of association $\widehat{MA}_1, \widehat{MA}_2, \ldots, \widehat{MA}_k$. If these estimates were free of systematic error, they would scatter randomly around the ''bull's-eye'' of the parameter. If they were free of random error, they would cluster tightly somewhere on the target. If they were free of both systematic error and random error, they would cluster tightly around the ''bull's-eye''—this is what the epidemiologist is ''shooting for'' (Figure 9.2).

A complementary way to conceptualize the problem of measurement error is to view each calculated estimate as the value of its underlying parameter plus ''error terms'' for random and systematic error:

$$\text{Observed estimate} = \text{Parameter} + \text{Random error} + \text{Systematic error}$$

The random and systematic errors inherent in an estimate bring the value of the estimate away from or toward the true value of the parameter. For example, an observed (calculated) risk ratio estimate of 3 might represent an overestimate of the risk ratio parameter by 1 with random error shifting the estimate up by 0.25 and systematic error shifting it up by an additional 0.75.

Although the amount of random error can be estimated from the data in the form of a standard error (or variance), the amount of systematic error cannot be easily quantified. Because random error and systematic error represent different types of problems with different types of solutions, they will be addressed separately.

9.2 Random error (imprecision)

Probability

It is often said that random error is governed by the laws of probability. However, *probability* is not easily defined when describing natural phenomena.

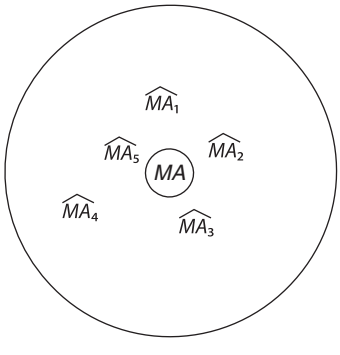


Figure 9.2 Target metaphor applied to measures of association estimates ($\widehat{MA}_i$) aiming at a parameter (MA).

A fundamental question that arises about probability and natural phenomena is whether we are talking about probability as an objective construct of the world, or whether we are discussing probability as a way to quantify our limited understanding of a situation. The former posits chance as an inherent property of nature that affects natural phenomena from the genes we inherit to the environment into which we are born. The latter posits probability against a background of incomplete knowledge.

Consider the flip a coin. We say that the probability it will turn up heads is 50%. An objective view of probability says that *if* the coin is flipped many, many times, we expect to see half of the flips turn up heads. It is easy to imagine that estimates of probabilities based on this method will become increasingly reliable as the number of replications increases. For example, if a coin is flipped 10 times, there is no guarantee that exactly 5 heads will be observed—the proportion of heads can range from 0 to 1, although in most cases we would expect it to be closer to 0.50 than to either 0 or 1. However, if the coin is flipped 100 times, chances are better that the proportion of heads will be close to 0.50. With 1000 flips, the proportion of heads will be an even better reflection of the true probability.

Now consider the flip of a coin in the context of our limited ability to predict its outcome. In theory, given enough information about the height and velocity of the flip, its rate of rotation, and the initial starting position of the coin in relation to the ground, the probability of it turning up a head might be predicted with greater than 50% certainty. Thus, the probability has changed based on what we know. According to this second view of probability, probabilities are defined in terms of the variability in the data that cannot otherwise be explained.

This second view of probability has relevance when studying disease occurrence. With no background knowledge, it might be sensible to say a person has less than a 1% probability of developing lung cancer during a lifetime. But if we then discover the person is male and smokes cigarettes, a better estimate for this probability might be 17% (Villeneuve and Mao, 1994). Thus, the objective state of affairs has not changed but the revised probability has changed because our knowledge about underlying conditions is now different. This revised probability conveys the extent to which we now believe the event is likely to occur.

Fortunately, appreciations of these two views of probability are both founded on our experience of the relative frequency of phenomena.

> When we say that the probability of a coin coming down heads on being tossed is one-half we have in mind, I think, that if it is tossed a large number of times it will come down heads in approximately half the cases. Even in extreme cases, say, when we attempt to assess the probability of a horse winning a given race, an event which cannot be repeated, we are, I think, picturing our estimation as one of a number of similar acts and assessing the relative frequency of the horse's victory in that population.''
>
> *Kendall (1947, pp. 165–166)*

Illustrative Example 9.1 Random noise in a laboratory experiment

Suppose a laboratory experiment wants to learn about the teratogenicity[a] of an agent. The investigator realizes that even genetically identical mice bred under identical laboratory conditions will express variable rates of congenital malformations when exposed to the teratogen. It is against this background of unexplained random noise upon which judgments will be made. The random noise in an experiment is given different names. It is called ''experimental error,'' ''biological variation,'' or ''chance.'' By any name, this is the variability in the outcome that cannot otherwise be explained.

In laboratory experiments we try to limit this random noise by controlling environmental conditions. But even here, unforeseen factors affect the outcome being studied. Ambient conditions vary, some mice eat more than others, batches of feed are not perfectly uniform, the pathologist diagnosing malformations may miss or misinterpret findings on necropsy, and so on. Randomness happens. Under an assumption of randomness, the effects from these unexplained factors are assumed to be independent of the treatment being studied and the random error associated with the estimate will follow a predictable probability distribution and can thus be dealt with mathematically.

Illustrative Example 9.2 Random and systematic error in a survey

Regardless of how one views probability and randomness, one thing is certain: as the size of a sample increases, the amount of random error associated with statistical estimation decreases. A brief consideration of sampling will illuminate this property.

Suppose an investigator wants to estimate the prevalence of smoking in a high school. The investigator is aware that any given random sample will not be an exact replica of the high school population. For example, a given sample of 10 may have three admitted smokers. However, the next sample of 10 may have five smokers. This is referred to as random sampling error.

The investigator is also aware that the amount of random sampling error in a sample will lessen if the sample size is increased. When based on a sample of, say, $n = 10$, sample-to-sample variability will be great. For example, a first sample may show 2/10 (20%), a second sample may show 5/10 (50%), and a third sample may show 3/10 (30%). However, larger samples of, say, $n = 100$, will derive more stable statistical results. For example, the first sample may show 29/100 (29%), the second sample 35/100 (35%), and the third sample 37/100 (37%). Estimates from large samples contain less random error than estimates from small samples.

On top of the problem of random sampling error, the investigator is concerned about systematic errors. Some of the teenagers in the survey may misrepresent their smoking habits (*information bias*). In addition, if given leeway, the interviewer may find it easier to select subjects from among teenagers who are easier or more pleasant to interview (*selection bias*). These systematic errors are quite different from the aforementioned random sampling error. Whereas the amount of random error will diminish with increasing sample size, the amount of systematic error is unaffected by sample size. The laws of probability are not suited for handling systematic errors.

Introduction to statistical inference

Statistical inference[b] is the process by which we address random error in data. The landmark statistical paper written by Gossett written under the pseudonym Student (1908) made the point in reference to the random error in experiments as follows:

> Any experiment may be regarded as forming an individual of a ''population'' of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

[a] Teratogenicity is the ability of an agent to cause malformations in developing embryos.
[b] Statistical inference is distinct from causal inference, as discussed in Section 4 of Chapter 2.

> Now any series of experiments is only of value in so far as it enables us to form a judgment as to *the statistical constants of the population* to which the experiments belong. (note emphasis added)

The same manner of thinking is applied to observational (nonexperimental) studies. To paraphrase ''Any series of observational studies is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the observational studies belong.'' Note that the statistical constants to which we refer are the parameters we have so far been discussing.

The two standard methods used to infer parameters in statistical analyses are estimation and hypothesis testing. The objective of **estimation** is to ''locate'' the value of the parameter. The objective of **hypothesis testing** is to test a claim about the parameter. Let us start by considering estimation.

Estimation (confidence intervals)

Estimation comes in two forms: point estimation and interval estimation. **Point estimation** provides the most likely value of the parameter. **Interval estimation** provides a range of values for the parameter in the form of a confidence interval.

Consider a rate difference of $\frac{5}{10\,000 \text{ person-years}}$. This is the **point estimate** for the true but unknown value of the underlying rate difference parameter. To construct a confidence interval for this parameter, we surround the point estimate with a calculated **margin of error**. Suppose the margin of error associated with the aforementioned risk difference estimate is $\frac{1.5}{10\,000 \text{ person-years}}$ with 95% confidence. Then, the 95% confidence interval for the rate difference parameter is:

$$\frac{5}{10\,000 \text{ person-years}} \pm \frac{1.5}{10\,000 \text{ person-years}}$$

$$= \left(\frac{3.5}{10\,000 \text{ person-years}} \text{ to } \frac{7.5}{10\,000 \text{ person-years}} \right)$$

In this example, $\frac{3.5}{10\,000 \text{ person-years}}$ is the **lower confidence limit** (**LCL**) and $\frac{7.5}{10\,000 \text{ person-years}}$ is the **upper confidence limit** (**UCL**) (Figure 9.3). We can now say with 95% confidence that the rate difference parameter after accounting for random error (but not for systematic error) lies between these limits.

The **confidence level** of a confidence interval quantifies our confidence in the procedure used to create the interval. Confidence intervals can be calculated at almost any confidence level. However, the most common level of confidence are 95, 90, and

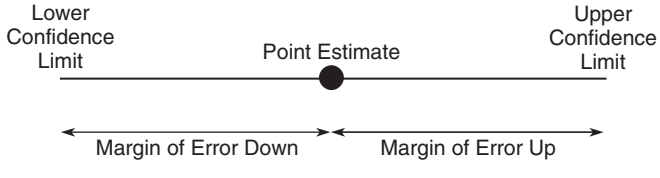


Figure 9.3 Representation of a confidence interval.

99%. A 95% confidence interval, for example, is designed to capture the parameter 95% of the time. This means that 5% of such intervals will miss the parameter. In addition, keep in mind that this technique addresses random error only and provides no protection against systematic errors.

The length of confidence interval reflects the estimate's **precision**. Long confidence intervals indicate that the estimate is imprecise; narrow confidence intervals indicate that the estimate is precise. All other things being equal, studies based on a large number of cases produce narrow (precise) confidence intervals; studies based on small numbers produce wide (imprecise) confidence intervals.

A common *mis*application of confidence intervals is to view them as ''significant'' or ''not significant'' by comparing their limits with a fixed value. This reduces estimation to a fixed-level hypothesis test—a process that is be discouraged when studying natural relationships. An example of this misapplication is illustrated by considering a relative risk of 1.7 with a 95% confidence interval of from 0.9 to 3.1. While some readers may interpret this confidence interval as insignificant because it does not rule out a relative risk of 1 with 95% confidence, it ignores the fact that data are as compatible with a relative risk of 3.1 as they are with a relative risk of 0.9. In addition, it is quite possible that, say, the 94% confidence for the same data would exclude a relative risk of 1 from its midst. 'Surely, God loves 94% confidence nearly as much as 95% confidence' (adaptation of a quote in Rosnow and Rosenthal, 1989).

Illustrative Example 9.3 Fat consumption and breast cancer

Figure 9.4 displays 95% confidence intervals for relative risks from 10 prospective cohort studies on total fat consumption and breast cancer. Study 2 and study 7 present the most precise estimates. Study 4 offers the least precise estimate. Some studies demonstrate relative risk point estimates that are slightly greater than 1 (e.g., studies 3, 4, 5, 6, and 9), and some demonstrate relative risks that are slightly less than 1 (e.g., studies 1, 2, and 7). None of the studies by themselves are precise enough to rule out chance as an explanation for their observed direction of the association. Thus, taken as a whole, these data suggest that there is no association between total fat intake and breast cancer risk.

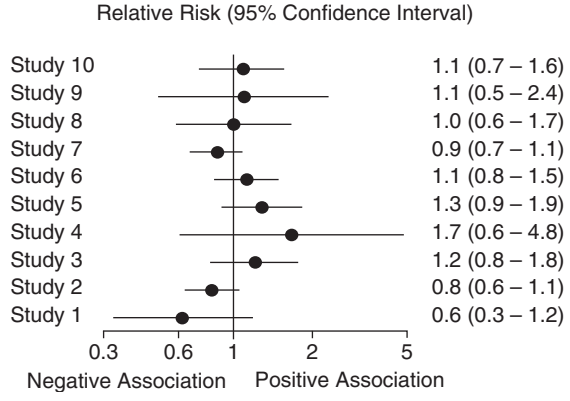


Figure 9.4 Confidence intervals for Illustrative Example 9.3. Relative risks of breast cancer and total fat intake, prospective cohort studies. Graph based on the data in Table 3 of Hunter and Willet (1994).

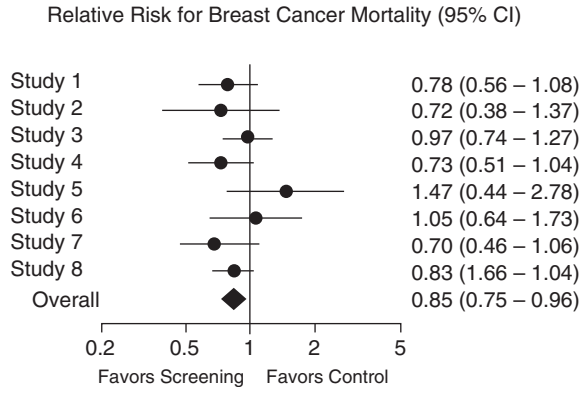


Figure 9.5 Relative risks of breast cancer mortality in relation to mammographic screening. (Based on data in Nelson *et al.*, 2009.)

Illustrative Example 9.4 Mammography for the prevention of breast cancer mortality

Figure 9.5 displays 95% confidence intervals for relative risks of death from breast cancer from eight different mammography screening trials for women aged 39–49. Five of the eight studies (studies 1, 2, 4, 7, and 8) show a negative association favoring screening, two show almost no association (studies 3 and 6), and one shows a positive association disfavoring mammographic screening (study 5). None of these studies by themselves are statistically significant. However, combining the results from all eight studies through a meta-analysis derived an overall RR of 0.85 (0.75–0.96), as shown by the diamond at the bottom of the plot.

Hypothesis testing (*p*-values)

Significance **hypothesis testing** (also called **significance testing**) is an elaborate process used to make judgments about statistical claims. Effective use of this technique requires a consideration of its underlying reasoning.

The test procedure begins by assuming no association between the study exposure and outcome. This assumption is called the **null hypothesis** ($\boldsymbol{H_0}$). It then specifies a probability model for prospective result that assumes that the null hypothesis is true. Based on this probability model, we are able to derive a measure of evidence called the *p*-value.[c] Generally, the smaller and smaller the *p*-value, the stronger and stronger the evidence is against the null hypothesis.

The underlying complexity behind the derivation of the *p*-value has made it susceptible to frequent **misinterpretation**. One common misinterpretation is to assume that the *p-value* represents ''the probability that null hypothesis is true.'' This

[c] The *p*-value is the probability of observing the data or data that are more extreme than the current data assuming the null hypothesis is true.

simply does not apply. In addition, *p-value*s are frequently misinterpreted as measures of practical significance, measures of the strength of an association, and measures of replication. Sadly, the *p-value* provides none of these results.

So how do we interpret *p-value*s in observational epidemiology? First, *p*-values must be viewed as continuous measure of evidence and *not* merely as ''significant'' or ''not significant.''[d] As noted, smaller and smaller *p*-values, provide increasing security that the observed association in the data cannot be simply ascribed to chance. Large *p*-values, say, greater than 0.10 or 0.15, indicate that we ought not pay too much attention to an observed association. Any *p*-values below this level begin to provide evidence that the *direction* of the observed association cannot be easily ascribed to chance and are thus ''significant.'' As the *p*-value gets smaller and smaller (say, 0.06 to 0.02 to 0.01), we get further reassurance that the observed direction of the association is not a ''random blip.'' Even more importantly, we must always interpret *p*-values in the context of the observed measure of association and other knowledge about the relationship being studied. Keep in mind that, like confidence intervals, *p*-values address random error only and provide no information about the systematic error in a study.

Illustrative Example 9.5 Childhood housing conditions and coronary heart disease mortality later in life

Table 9.1 lists relative risks relating childhood housing conditions and coronary heart disease mortality later in life from a study by Dedman and coworkers (2001). The relatively high *p*-values associated with crowding, type of toilet, ventilation, and cleanliness suggests that chance (random error) is a plausible explanation for the observed association. In contrast, the lack of any indoor tap water was associated with a relative risk of 1.7 and *p*-value of 0.01, suggesting that this association cannot be easily explained by chance.

9.3 Systematic error (bias)

The second and perhaps more pernicious form of error in epidemiologic research is **systematic error**. ''Systematic error'' is also called ''bias.''

Bias is defined as the difference between the statistical expectation of an estimate and the parameter it purports to estimate. If we let θ represent a generic parameter, $\hat{\theta}$ represent its statistical estimator, and $E(\hat{\theta})$ represents the expected value of the estimator, then an estimate is unbiased when $E(\hat{\theta}) = \theta$ and is biased when $E(\hat{\theta}) \neq \theta$. When $E(\hat{\theta}) > \theta$, a positive bias exists. When $E(\hat{\theta}) < \theta$, a negative bias exists. Bias thus has a direction (i.e., positive bias or negative bias) and an amount (i.e., a lot of bias or a little bias).

It is helpful to identify these three broad classes of bias:

- **Selection bias**, due to the manner in which subjects are selected for study.

[d] *p*-values should be reported with two significant digit accuracy (e.g., $p = 0.028$) and not as an inequality (*not* $p < 0.05$).

Table 9.1 Relative risks of coronary heart disease based on childhood household conditions (Dedman *et al*., 2001; Galobardes *et al*., 2004).

Household condition	Relative risks (95% CI)[a]		*p*-value (trend)
Crowding	**Persons per room**	***RR***	$p = 0.15$
	<1.5	0.7 (0.5, 1.1)	
	1.5–<2.5	1.0 (ref.)	
	2.5–<3.5	1.1 (0.8, 1.7)	
	≥3.5	1.2 (0.7, 1.9)	
Tap water	No vs. yes	1.7 (1.1, 2.6)	$p = 0.01$
Type of toilet	Flush inside, not shared	0.7 (0.5, 1.2)	$p = 0.14$
	Flush outside, shared	1.0 (ref.)	
	No flush	0.7 (0.4, 1.4)	
Ventilation	Very good	1.0 (ref.)	$p = 0.24$
	Fair	1.1 (0.8, 1.5)	
	Poor	1.4 (0.8, 2.3)	
Cleanliness	Very good	1.3 (0.8, 2.1)	
	Fair	1.0 (ref.)	
	Poor	1.4 (1.0, 2.0)	$p = 0.21$

[a] Adjusted for income, food expenditures, Townsend deprivation score, and childhood social class.

- **Information bias**, due to the manner in which information used in the study is erroneously measures or classified.
- **Confounding**, due to the influence of extraneous factors on observed associations.[e]

Selection bias

Selection bias refers to a distortion in a statistical estimate resulting from the manner in which subjects are selected for study. In studies of causal effects, this type of bias comes about when the association between the exposure and disease differs between those who participate in the study and those who do not.

Many instances of selection bias have been documented. Some of the more colorful examples come from faulty political polls from the middle of the 20th century. Darrell Huff in his lucid and amusing book *How to Lie with Statistics* (1954) tells how pre-election polls of the 1936 election routinely had President Franklin Delano Roosevelt losing to his opponent Alf Landon by wide margins. These erroneous predictions were based on telephone polls at a time when many people did not have phones in their homes. Telephone samples, therefore, did not represent the general electorate: phone

[e] Some epidemiologists consider confounding separately from other sources of bias because it is strongly linked to subject matter, and reserve the term bias for errors on the part of the study. However, this and many other texts considers confounding as a distinct form of bias.

owners were economically better off than non-owners and were more likely to vote Republican. In effect, the sample elected Landon, but the electorate elected Roosevelt.

Another political survey mishap occurred during the 1948 re-election bid by President Harry S. Truman. Many of the major political polling organizations of the time had Truman's Republican rival, Thomas E. Dewey, slated to win the election. Apparently, data collectors for these polls had some freedom to choose whom they interviewed, and Republicans were typically easier to interview or more approachable than Democrats (Mosteller, 1949). These political polls illustrate the need to take care in sampling.

Examples of specific types of selection biases in epidemiologic research include:

1 **Hospital admission bias**, also called **Berkson's bias** (1946), is a form of selection bias that affects epidemiologic studies done in hospitalized populations. When hospitalization rates differ for different exposure groups, the relation between an exposure and disease in the hospital will not reflect the relation in the population that serves as the source of cases.
2 **Prevalence-incidence bias** (Neyman, 1955) occurs when prevalent cases are used to study exposure–disease relations. This bias is related to the fact that prevalent cases represent survivors of the condition being studied. Since survivors may be atypical, a study based on prevalent cases may misrepresent the relationship between the exposure and disease. In studying disease etiology, it is almost always preferable to use incident cases.
3 **Nonresponse bias** and **withdrawal bias** are forms of selection bias that occur when study subjects refuse to participate in a study and when current study participants selectively drop out of a study once the study has begun. ''Refusers'' and ''discontinuers'' often differ systematically from people who are compliant study subjects.
4 **Publicity bias** occurs when media or medical attention increases awareness of a perceived health problem in some groups more than others. For example, if a celebrity publicizes a putative cause for a particular illness, this stimulates individuals to wonder if they or their children might have developed the same illness due to the imagined cause. This phenomenon occurred widely in the wake of the Jenny McCarthy campaign to blame the MMR (measles, mumps, rubella) vaccine as the cause of autism in her child. Ms. McCarthy appeared on multiple media outlets in the United States (e.g., The Oprah Winfrey Show, Larry King Live) supporting her view with flawed fraudulent studies (Godlee *et al*., 2011).
5 **The healthy worker effect** is a form of selection bias that occurs because seriously ill people are often unable to join or remain in the workforce. This bias expresses itself as lower than expected morbidity and mortality rates among workers. Consequently, comparisons of worker groups with the general population tends to understate risks associated with work-related exposures. One way to avoid this type of bias is to compare workers in an exposed job category to another group of workers in a nonexposed job category. Another method to mitigate this type of bias is to use a ''length of history screen'' in which a certain period of time must elapse before the worker is enrolled in a study cohort.

Information bias

Information bias is a distortion in a measure of association or effect due to measurement error or misclassification of subjects on one or more study variables. This type of bias arises from measurement device defects, questionnaire and interview procedures that do not measure what they purport to measure, inaccurate diagnostic procedures, incomplete or incorrect data sources, and data processing errors. When the information being collected is categorical, this type of bias is referred to **misclassification bias**.

In addressing misclassification, we must consider whether such misclassifications are nondifferential or differential. **Nondifferential misclassification** refers to misclassification that occurs equally in the groups that are being compared. **Differential misclassification** refers to misclassification that occurs unequally in groups. This distinction has important implications because nondifferential misclassification will tend to bias measures of association either toward the null[f] or not at all (Bross, 1954; Poole, 1985). Numerical illustrations will make this point.

Table 9.2 presents hypothetical examples of nondifferential misclassification demonstrating no bias and bias toward the null. Table 9.2A presents the data as it should be, without misclassification, demonstrating a risk ratio of 2.00 and an odds ratio

Table 9.2 Effects of nondifferential misclassification, examples.

9.2A Accurate data

	D+	D–	N
Exposed	100	900	1000
Nonexposed	50	950	1000

$RR = \frac{100/1000}{50/1000} = 2.00$ $\quad OR = \frac{100 \cdot 950}{50 \cdot 900} = 2.11$

9.2B 10% Nondifferential misclassification of case status

	D+	D–	N
Exposed	90	910	1000
Nonexposed	45	955	1000

$RR = \frac{90/1000}{45/1000} = 2.00$ (no bias) $\quad OR = \frac{90 \cdot 955}{45 \cdot 910} = 2.10$ (biased minimally toward null)

9.2C 10% Nondifferential misclassification of case status and 20% nondifferential misclassification of exposure status

	D+	D–	N
Exposed	72	728	800
Nonexposed	63	1137	1200

$RR = \frac{72/800}{63/1137} = 1.71$ (biased toward null) $\quad OR = \frac{72 \cdot 1137}{63 \cdot 728} = 1.78$ (biased toward null)

[f] A bias toward null means that the observed estimate will underestimate the positive or negative effects of the exposure.

of 2.11. Table 9.2B presents the data in which 10% of the cases have been nondifferentially misclassified as noncases in both the exposed and nonexposed series. The risk ratio under this scenario is still 2.00 (unbiased), while the odds ratio is now 2.10 (biased slightly toward the null). Table 9.2C presents the data with a 10% nondifferential misclassification of case status and 20% nondifferential misclassification of the exposure status. Under these conditions, the risk ratio is now 1.71 and odds ratio is 1.78 (both biased toward the null). In effect, nondifferential misclassification cannot explain apparent associations. In contrast, differential misclassification can create biases both toward the null and away from the null, resulting in artificial association and hiding real ones. Examples of specific types of information biases follow.

1 **Recall bias** refers to the form of information bias that ensues when inaccurate information is reported by study subjects or their surrogates. For example, it is not uncommon for cases in case–control studies to scrutinize their histories with more rigor and imagination than controls. When this occurs, an odds ratio indicating an increased risk will be biased away from the null.
2 **Diagnostic suspicion bias** refers to the type of bias that occurs when the exposed group undergoes greater diagnostic scrutiny than the nonexposed group (or vice versa). Horowitz and Feinstein (1978) suggested that this type of bias was operative in a case–control study of estrogen and endometrial cancer in which increased diagnostic attention was directed toward women taking hormone replacement therapy. The authors suggested that estrogen-induced bleeding may have invoked referral for diagnosis of endometrial cancer symptoms that may otherwise have gone undetected. Thus, the estrogen exposure might have led to earlier detection of uterine cancer without actually increasing risk. (Unopposed estrogen has since been confirmed as a risk factor for endometrial cancer.)
3 **The Clever Hans effect** (**obsequiousness bias**) occurs when subjects alter their responses in the direction they perceive to be desired by the investigator. This bias is named after a trained horse that could apparently do simple mathematics. Although never proved, it is presumed the clever equine Hans picked up nonverbal clues from his trainer that helped him determine when to stop stomping his hoof in response to numerical questions. Like Hans' trainer, interviewers may send nonverbal clues to study subjects, thus influencing responses.

Confounding

Confounding (from the Latin *confundere*, to mix together) occurs when an association between a study exposure and study outcome is brought about by the influence of an extraneous factor (or extraneous factors) lurking in the background. This form of bias derives from inherent difference in risk between the exposed group and nonexposed group that would exist even if the exposure were absent. Extraneous factors that cause the imbalance are called **confounders**.

Confounders have the followed **properties** (Figure 9.6):

1 They are associated with the exposure.
2 They are independent risk factors for the study outcome.
3 They are *not* intermediate in the causal pathway and are not a consequence of the disease (Greenland and Robins, 1986).

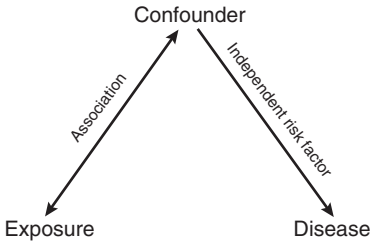


Figure 9.6 Properties of a confounder.

Illustrative Example 9.6 Alcohol and lung cancer, confounding by cigarette smoking

As an example of confounding, consider smoking as a potential confounder when studying the relationship between alcohol consumption (the exposure) and lung cancer (the study outcome). Since (1) frequent alcohol consumers are more likely to smoke, (2) cigarette use is an independent risk factor for lung cancer, and (3) smoking is not in the causal pathway between alcohol use and lung cancer, smoking is likely to confound the association between alcohol and lung cancer (Figure 9.7).

Table 9.3 presents fictitious data demonstrating an overall risk ratio of 3 for alcohol consumption and lung cancer. However, when data are stratified into nonsmokers and smokers, nonsmokers demonstrate a risk ratio of 1. Within smokers, the risk ratio is also 1. The reason for this paradoxical finding is that 80% of the alcohol consumers were smokers, while only 10% of those who did not consume alcohol were smokers. This allowed smoking to confound the association between alcohol consumption and lung cancer.

Illustrative Example 9.7 Helicopter evacuation and survival following motor vehicle accidents, confounding by severity of injury

Injuries sustained in motor vehicle accidents are normally transported to the hospital by ambulance. However, when the accident is severe or requires special treatment, the evacuation may be accomplished by helicopter. Thus, the association between helicopter evacuation (the exposure) and surviving a traffic accident (the outcome) may be confounded by the severity of the injury (the confounder). Figure 9.8 shows how the properties of confounding have been satisfied. Table 9.4 presents a numerical illustration from Oppe and De Charro (2001) that demonstrates the phenomenon. Notice that the positive association between helicopter transportation and death is reversed after considering the seriousness of the injuries. Reversal of associations by confounding factors is referred to as Simpson's Paradox (Simpson, 1951).

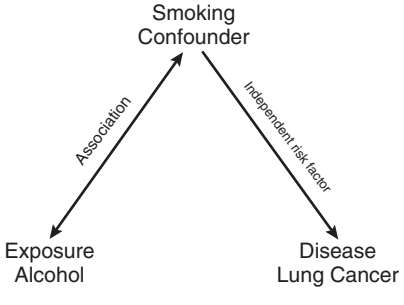


Figure 9.7 Illustrative Example 9.6, smoking confounds the observed relationship between alcohol and lung cancer.

Table 9.3 Numerical examples for Illustrative Example 9.6.

Smokers and nonsmokers combined

	LungCA +	LungCA –	
Alcohol +	84	99 916	100 000
Alcohol –	28	99 972	100 000

$$RR = \frac{84/100\,000}{28/100\,000} = 3.00$$

	Nonsmokers			Smokers		
	LungCA +	LungCA –		LungCA +	LungCA –	
Alcohol +	4	19 996	20,000	80	79 920	80 000
Alcohol –	18	89 982	90 000	10	9 990	10 000

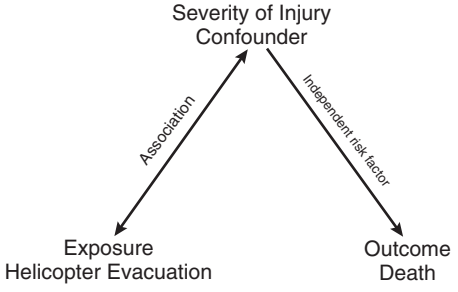


Figure 9.8 Illustrative Example 9.7, severity of injury confounds the observed relationship between helicopter evacuation and survival.

Table 9.4 Numerical examples for Illustrative Example 9.7.

All injuries

	Died	Survived	N
Helicopter	64	136	200
Road	260	840	1100

$$RR = \frac{64/200}{260/1100} = \frac{0.3200}{0.2364} = 1.35$$

Serious injuries

	Died	Survived	N
Helicopter	48	52	100
Road	60	40	100

$$RR = \frac{48/100}{60/100} = 0.80$$

Less serious injuries

Died	Survived	N
16	84	100
200	800	1000

$$RR = \frac{16/100}{200/100} = 0.80$$

Let us conclude this section with a real example in which the unraveling of a confounded association changed our beliefs about whether menopausal women should or should not routinely take hormone replacement therapy.

Illustrative Example 9.8 Hormone replacement therapy

The use of supplementary estrogen and progestin (hormone replacement therapy, HRT) had become commonplace for women at around the time of menopause by the end of the 20th century. One of the reasons for the widespread use of HRT was the belief, supported by observational studies, that HRT lowered rates of cardiovascular disease and death in women. In 1998, a randomized trial of HRT in women who had already had heart disease found no benefit to hormone use (Hully *et al*., 1998). Then, in 2002, an experimental study initially discussed in Illustrative Example 6.1 known as the Women's Health Initiative (WHI, 2002) showed that indiscriminant use of HRT actually increased the risk of cardiovascular disease and death in women. Several types of confounding were proposed to help explain the contradictory results of prior observational studies with the WHI trial.

One explanation, dubbed ''the healthy user effect'' suggested that women who seek treatments perceived as being health-promoting, such as HRT, also tended to engage in other activities intended to improve health. This explanation was supported by the observation that women who chose to take HRT tended to be wealthier, have lower body weight, exercise more often, and have fewer risk factors for cardiovascular disease than those women who did not use HRT (see Illustrative Example 7.5). Thus, exposure to HRT may have been a marker for health consciousness rather than a protective factor for cardiovascular disease.

Another explanation for the apparently protective cardiovascular effects in earlier observation studies involved confounding by socioeconomic status. Support for this explanation comes from the fact that the apparent protective effects of HRT in observational studies were primarily restricted to those studies that did not adequately control for socioeconomic status, while observational studies that did adequately control for socioeconomic status tended to find no protective effect (Figure 9.9). It is plausible that higher socioeconomic status was associated with both more frequent HRT use and a lower baseline risk of CHD, confounding the observed association in some of the prior studies on HRT and CHD.

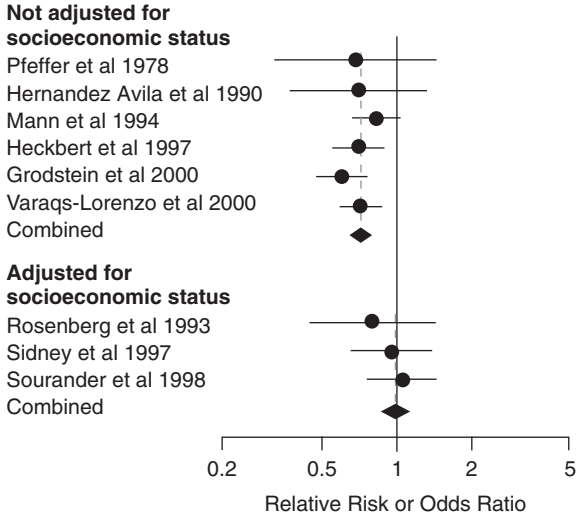


Figure 9.9 Observational studies of HRT and coronary heart disease. SES may have confounded the studies showing protective cardiovascular effects. See Illustrative Example 9.8 for further explanation. (Adapted from Humphrey *et al.,* 2002.)

Exercises

9.1 Biased prevalences. For each survey described below, predict whether the prevalence estimate will be affected by selection bias or information bias. In addition, predict whether the bias will inflate or deflate the true prevalence.

(A) A survey of venereal disease based on self-reporting.

(B) A survey of disability in the elderly based on a sample of senior citizens attending dance lessons.

(C) A survey of coronary artery disease based on the question ''Has a doctor ever told you that you have coronary artery disease?''

(D) A survey of self-reported smoking in teenagers.

(E) A survey of carpal tunnel syndrome in employees who are *not* eligible for disability compensation.

(F) A telephone survey to determine the prevalence of senile dementia.

9.2 Race, socioeconomic status, and high blood pressure. When an association is found between two factors, there is always the possibility that the association is due to an extraneous factor. This phenomenon is known as confounding. Stratifying on the extraneous factor is useful in assessing whether the observed association is due to the confounding effects of the extraneous factor. Suppose, for example, we discover an association between low socioeconomic status and hypertension: the prevalence of hypertension in low-, intermediate-, and high-socioeconomic 50–54 year old men is found to be 33, 21, and 17%, respectively. However, the literature on hypertension research reveals that 50–54 year old African–American men have higher blood pressures on the average than that of other racial-ethnic groups. We also know that race and socioeconomic status are associated.

(A) Is it likely that the association between socioeconomic status and hypertension is confounded by race? Why?

(B) Now suppose we stratify the data according to race and find that within Caucasians the prevalence of hypertension is 17, 16, and 16% in low-, intermediate-, and high-socioeconomic groups. In African–Americans we find prevalence of 29, 28, and 27%, respectively. Based on these subgroup analyses, what do you conclude?

9.3 Meta-analysis of coronary heart disease treatment options. A meta-analysis by Hlatky and coworkers (2009) compared coronary artery bypass graft (CABG) with percutaneous coronary interventions (PCI) for the treatment of multi-vessel coronary disease by pooling individual data on 7812 patients from ten randomised trials.

(A) The article states ''observational studies [on CABG and PCI] have been confounded by treatment selection bias.'' It is reasonable to assume that in clinical practice the more severe coronary artery blockages might be channeled to CABG. Based on your understanding of the properties of confounding, discuss specifically how treatment selection bias would confound the results of observational studies comparing CABG and PCI.

(B) In an observational study of CABG vs. PCI, would you expect treatment selection bias to favor CABG or PCI? Explain your reasoning.

(C) Explain how randomization mitigates the potential for treatment selection bias.

(D) The authors of the article go on to state that the ''Overall mortality was similar in the treatment groups; 574 of the 3889 patients died in the CABG group compared with 628 of the 3923 patients in the PCI group,'' $p = 0.12$. Calculate the risks of death in each group and interpret the reported p value in this context.

(E) Figure 9.10 compares mortality after coronary artery bypass graft (CABG) or percutaneous coronary intervention (PCI) in selected subgroups. Based on these findings, are there any subgroups that demonstrate a survival advantage of one procedure or the other? Explain your response.
Note: Finding different effects in subgroups is called **effect measure modification** or **statistical interaction**. Chapter 15 introduces methods to detect and address statistical interactions.

9.4 Alternative explanations for high rates of disability in monks. It is said that any good epidemiologist is able to propose alternative noncausal explanations for observed associations. An epidemiologic study based on a self-administered questionnaire finds a higher rate of disabilities related to activities of daily living (e.g., difficulty sitting down or getting up from a chair) in Dutch Trappist and Benedictine monks than in other comparably aged Dutch men (Mackenbach *et al.*,

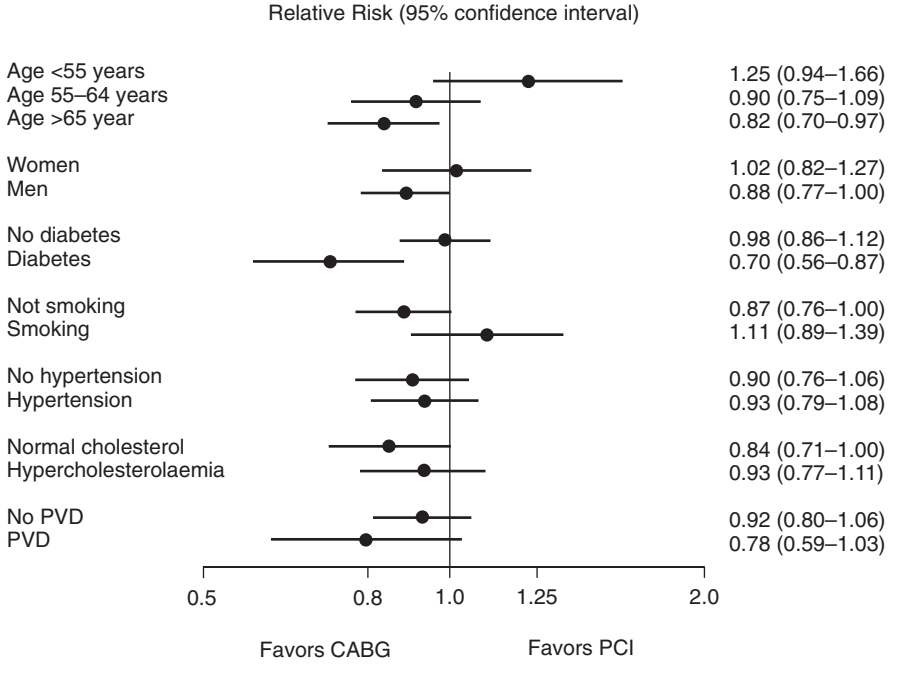


Figure 9.10 Mortality after treatment with coronary artery bypass graft (CABG) versus percutaneous coronary intervention (PCI), selected subgroups. (Based on data in Hlatky *et al.*, 2009.)

1993). Come up with explanations for these results that are unrelated to the daily activities of being a monk.

Review questions

R.9.1 Name the two basic types of error in epidemiologic research.

R.9.2 How do *parameters* differ from *estimates*?

R.9.3 Provide a synonym for *systematic error*.

R.9.4 Provide a synonym for *random error*.

R.9.5 Provide an antonym for *biased*.

R.9.6 Provide an antonym for *precise*.

R.9.7 List ways in which random error differs from systematic error.

R.9.8 True or false? Probability models are used to adjust for bias in epidemiologic studies. Explain your response.

R.9.9 What are the two common forms of statistical inference?

R.9.10 Suppose the sample size of an observation study could be expanded to be infinitely large. How would this affect the amount of random error in the study? How would it affect systematic error?

R.9.11 Name the three general categories of bias in epidemiologic research.

R.9.12 What is the effect of nondifferential misclassification on measures of association?

R.9.13 True or false? A bias toward null means that the observed measure of association will underestimate the true risks or benefits associated with exposure.

R.9.14 Define *confounding*.

R.9.15 List the properties of a confounder.

R.9.16 What does the Latin word *confundere* mean?

R.9.17 Use of hospitalized cases and controls in case–control studies could result in what type of bias?

R.9.18 Suppose the cases in a case–control study gave more complete responses about exposures to potential risk factors than the controls. What type of bias will this cause? Will the odds ratio in the study be biased toward the null or away from the null?

R.9.19 Suppose the code book for a data set gets mixed-up so that exposed individuals are mistakenly coded as nonexposed and vice versa. What type of bias will this cause?

R.9.20 Not all risk factors are confounders. Why?

R.9.21 It is *not* always essential for studies of causal factors to be done in populations that are demographically representative of other populations. For example, the results of a study on smoking and lung cancer in men can be generalized to the effects of smoking on lung cancer in women. Why do you suppose this is the case?

R.9.22 True or false? Confidence intervals adjust for both random and systematic sources of error providing confidence in the study's results.

References

Berkson, J. (1946) Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, **2**, 47–53.

Bross, I.D.J. (1954) Misclassification in 2 × 2 tables. *Biometrics*, **10**, 478–486.

Dedman, D.J., Gunnell, D., Davey Smith, G., and Frankel, S. (2001) Childhood housing conditions and later mortality in the Boyd Orr cohort. *Journal of Epidemiology and Community Health*, **55** (1), 10–15.

Galobardes, B., Lynch, J. W., and Davey Smith, G. (2004) Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. *Epidemiologic Reviews*, **26**, 7–21.

Godlee, F., Smith, J., and Marcovitch, H. (2011) Wakefield's article linking MMR vaccine and autism was fraudulent. *British Medical Journal*, **342**, c7452.

Greenland, S., and Robins, J.M. (1986) Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, **15**, 413–419.

Hlatky, M.A., Boothroyd, D.B., Bravata, D.M., Boersma, E., Booth, J., Brooks, M.M., *et al*. (2009) Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials. *Lancet*, **373** (9670), 1190–1197.

Horwitz, R.I., and Feinstein, A.R. (1978) Alternative analytic methods for case–control studies of estrogens and endometrial cancer. *New England Journal of Medicine*, **299**, 1089–1094.

Huff, D. (1954) *How to Lie with Statistics*, W.W. Norton, New York.

Humphrey, L.L., Chan, B.K., and Sox, H.C. (2002) Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Annals of Internal Medicine*, **137** (4), 273–284.

Hunter, D.J., and Willett, W.C. (1994) Diet, body build, and breast cancer. *Annual Review of Nutrition*, **14**, 393–418.

Hulley, S., Grady, D., Bush, T., Furberg, C., Herrington, D., Riggs, B., *et al*. (1998) Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA*, **280** (7), 605–613.

Kendall, M.G. (1947) *The Advanced Theory of Statistics*, 3rd edn, vol. I, C. Griffin & Company, London.

Mackenbach, J.P., Kunst, A.E., de Vrij, J.H., and van Meel, D. (1993) Self-reported morbidity and disability among Trappist and Benedictine monks. *American Journal of Epidemiology*, **138** (8), 569–573.

Mosteller, F. (1949) *The Pre-election Polls of 1948*, Social Science Research Council, New York.

Nelson, H.D., Tyne, K., Naik, A., Bougatsos, C., Chan, B.K., and Humphrey, L. (2009) Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, **151** (10), 727–737, W237–W742.

Neyman, J. (1955) Statistics—servant of all sciences. *Science*, **122**, 401–406.

Oppe, S., and De Charro, F.T. (2001) The effect of medical care by a helicopter trauma team on the probability of survival and the quality of life of hospitalised victims. *Accident Analysis & Prevention*, **33** (1), 129–138.

Poole, C. (1985) Exception to the rule about nondifferential misclassification. (Abstract). *American Journal of Epidemiology*, **122**, 508.

Rosnow, R.L., and Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, **44** (10), 1276–1286.

Simpson, E.H. (1951) The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B*, **13** (2), 238–241.

Student (1908) The probable error of a mean. *Biometrika*, VI, 1–25.

Tagore, R. (1916) *Stray birds* (Web). Macmillan Company, New York. Available: *http:// 209.11.144.65/eldritchpress/rt/stray.htm* (accessed 1 August 2002).

Villeneuve, P.J., and Mao, Y. (1994) Lifetime probability of developing lung cancer, by smoking status, Canada. *Canadian Journal of Public Health*, **85** (6), 385–388.

WHI (2002) Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*, **288** (3), 321–333.

CHAPTER 10
Screening for Disease

10.1 Introduction

10.2 Reliability (agreement)
- Essential background
- The kappa statistic
- The kappa paradox

10.3 Validity
- Sensitivity and specificity
- Predictive value positive and predictive value negative
- True prevalence and apparent prevalence
- Relation between prevalence and the predictive value of a positive test
  - Bayesian formulas for predictive value
- Relation between prevalence and the predictive value of a negative test
- Selecting a cutoff point for positive and negative test results

Summary
- Key points
- Reliability notation
- Validity notation

Exercises

Review questions

References

10.4 Chapter addendum (case study)
- Screening for antibodies to the human immunodeficiency virus
  - Objectives
  - Part I
  - Part II
  - Part III
  - Part IV

Further reading—screening for HIV

Further reading—general concepts of screening
- Answers to case study: screening for antibodies to the human immunodeficiency virus
  - Part I
  - Part II
  - Part III

Certainty of classification is a will-o′-wisp hovering over a deep quicksand.

John Tukey (1969, p. 86)

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

10.1 Introduction

This chapter considers the accuracy of diagnostic tests and procedures. It also considers implications of diagnostic test accuracy in population-based screening programs.

The accuracy of a diagnostic test is a function of the procedure and technology used to collect information. Data can be derived by personal interview, self-administered questionnaire, abstraction of medical records, or direct examination of study subjects. Direct examinations may be based on symptoms, signs, and diagnostic test results. **Symptoms** are subjective sensations, perceptions, and observations made by the patient. Examples of symptoms are pain, nausea, fatigue, and dizziness. **Signs** are perceptions and observations made by an examiner. Although signs tend to be more objective than symptoms, they are still influenced by the skill and judgment of the examiner. **Diagnostic tests** are measures of physical, physiologic, immunologic, and biochemical processes. Tests can range from the mundane (e.g., body temperature) to the technical (e.g., clinical chemistry). It is important to note that different methods of case ascertainment may derive different epidemiologic results (Table 10.1).

Even objective procedures demonstrate intra- and inter-observer variability. Figure 10.1 displays blood glucose determinations on a single pooled blood specimen sent to ten different clinical laboratories. Values derived by each lab were compared with the true glucose level determined by a definitive (''gold-standard'') technique with no known sources of error, that is, isotope dilution-mass spectrometry. The true value determined by this state-of-the-art method was 5.79mmol/l. However, readings within clinical labs (intra-observer reliability) and between clinical labs (inter-observer reliability) varied widely.

The accuracy of any diagnostic method is characterized by two distinct elements: its **reliability** (agreement upon repetition) and its **validity** (ability to discriminate between people with and without disease). These elements are considered separately.

Table 10.1 Comparison of prevalence estimates of selected chronic conditions as determined by household interviews and clinical evaluations, all ages combined.

Condition	Prevalence per 1000	
	Household interview	Clinical evaluation
Heart disease	25	96
Hypertension	36	117
Arthritis (any type)	47	75
Neoplasms (any type)	8	55

Source: Adapted from Lilienfeld and Lilienfeld (1980, p. 150); data from Commission on Chronic Illness (1957).

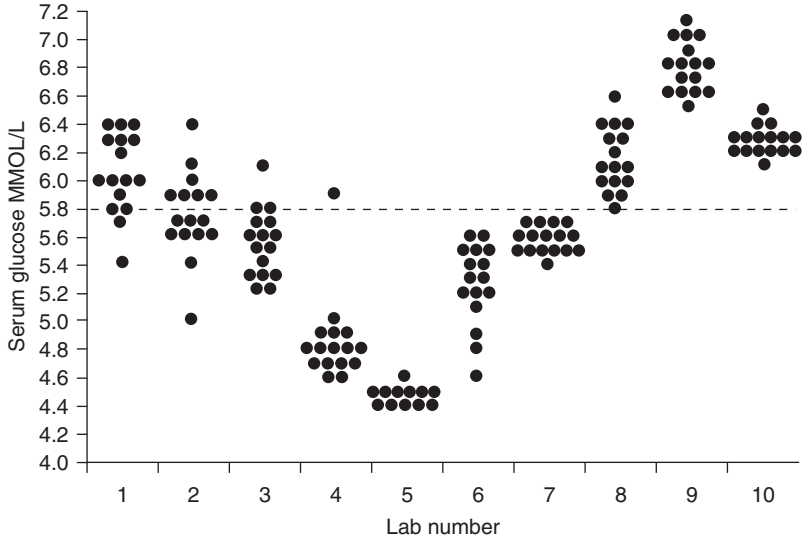


Figure 10.1 Blood glucose determinations of a pooled sample of blood according to ten clinical laboratories in Sweden. The horizontal dashed line represents the actual glucose level of the samples as determined by a definitive method known as isotope dilution-mass spectrometry (Based on data in Björkhem *et al*., 1981 and Ahlbom and Norell, 1990, p. 17).

10.2 Reliability (agreement)

Essential background

Reliability refers to the extent to which intra- or inter-rater ratings agree from one evaluation to the next. Thus, this parameter is also referred to as **agreement** and **reproducibility**.

Measurements that fail to agree with each other upon repetition are unreliable, whereas those with high levels of agreement are reliable. For example, if two physicians consistently agreed with each other on the diagnoses of a series of patients, this would indicate a high degree of inter-rater reliability. In contrast, if there were many diagnostic disagreements, this would indicate inter-rater unreliability.

A classic 1966 study of diagnostic reproducibility by Lilienfeld and Kordan found a significant number of discrepancies in the interpretation of chest X-rays read by radiologists. Using six diagnostic categories, the observed level of diagnostic agreement was a modest 65.1% (Table 10.2). When the diagnostic classification scheme was simplified to form only two diagnostic categories—significant pulmonary lesion, yes or no—diagnostic agreement improved to 89.4% (Table 10.3). These agreement levels are all the less impressive when one considers that among the 3558 X-rays labeled as positive by at least one of the radiologists, agreement was present in only 1467 (41.2%)—and this does not account for agreement due to chance.

Proportion of agreement in subjects labeled positive by at least one radiologist

$$= \frac{\text{total no. of agreements}}{\text{total no. X-rays with at least one positive diagnosis}} = \frac{1467}{1467 + 1309 + 782} = 0.412$$

Table 10.2 Comparison of two different radiologists' interpretations of chest X-ray films; outlined diagonal represents areas of diagnostic agreement.

	Radiologist B[a]						
Radiologist A	SN	OSPA	CV	NSA	NEG	TU	Total
SN	61	16	1	9	8	0	95
OSPA	70	1320	63	861	367	33	2714
CV	19	151	1322	369	1880	62	3803
NSA	25	407	43	1716	1656	40	3887
NEG	28	157	91	680	8475	50	9481
TU	0	2	0	4	47	0	53
Total	203	2053	1,520	3639	12 433	185	20 033

Source: Lilienfeld and Kordan (1966, p. 2147).

[a] *Abbreviations:* SN = suspect neoplasm; OSPA = other significant pulmonary abnormality; CV = cardiovascular abnormality; NSA = nonsignificant abnormality; NEG = negative; TU = technically unsatisfactory X-ray film.

$$\text{Observed proportion of agreement } (p_o) = \frac{\text{total no. of agreements}}{\text{total no. of readings}} = \frac{61 + 1320 + 1322 + 1716 + 8475}{20\,333 - (185 + 53)}$$

$$= 0.651$$

Table 10.3 Comparison of two different radiologists' interpretations of chest X-ray films.[a]

	Radiologist B[a]		
Radiologist A	+	–	Total satisfactory films
+	1467	1309	2776
–	782	16 232	17 014
Total satisfactory films	2249	17 541	19 790

Source: Lilienfeld and Kordan (1966, p. 2147).

[a] + = Suspected neoplasm or other significant pulmonary abnormality.
– = Cardiovascular abnormality, nonsignificant abnormality, negative.
X-ray films that were deemed to be technically unsatisfactory by either reviewer were eliminated from this analysis.

$$\text{Observed proportion of agreement } (p_o) = \frac{1467 + 16\,232}{19\,790} = 0.894$$

The kappa statistic

The **kappa statistic (κ)** was developed to measure the level of agreement between rates that occurs beyond that due to chance (Cohen, 1960). Consider an experiment that simultaneously flips two coins (Figure 10.2). We expect that the two coins would agree heads or tails half of the time. Thus, the expected level of agreement due to chance is 50%. The kappa statistic is constructed so that when the observed agreement is no greater than that which is expected due to chance, $\kappa = 0$. Greater than chance agreement leads to positive values of κ. When there is complete agreement, $\kappa = +1$. One widely used benchmark scale for characterizing the strength of agreement indicated by kappa values is shown as Table 10.4.

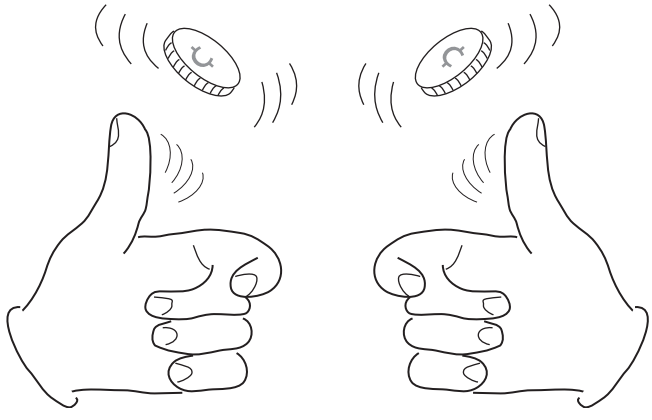


Figure 10.2 Some agreements are due to chance.

Table 10.4 Benchmark scale for interpreting kappa according to Landis and Koch (1977).

Kappa statistic	Strength of agreement
<0.0	Poor
0.0–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

Table 10.5 Notation for measuring agreement for a binary diagnostic test.

	Rater B		
Rater A	+	–	
+	a	b	g_1
–	c	d	g_2
	f_1	f_2	N

To calculate kappa for a binary outcome (condition present or absent), data are laid out in a two-by-two table with the notation shown in Table 10.5. Using this notation, the **observed proportion of agreement** is

$$p_{\text{obs}} = \frac{a+d}{N} \tag{10.1}$$

the **expected proportion of agreement due to chance** is

$$p_{\text{exp}} = \frac{f_1 g_1 + f_2 g_2}{N^2} \tag{10.2}$$

and Cohen's **kappa statistic** is

$$\kappa = \frac{p_{\text{o}} - p_{\text{e}}}{1 - p_{\text{e}}} \tag{10.3}$$

Illustrative Example 10.1 Kappa statistic

For the X-ray inter-rater agreement data presented in Table 10.3:

$p_{\text{obs}} = \frac{a+d}{N} = \frac{1467 + 16232}{19790} = 0.8943 \rightarrow$ the observed level of agreement is 89.4%

$p_{\text{exp}} = \frac{f_1 g_1 + f_2 g_2}{N^2} = \frac{(2249 \cdot 2776) + (17541 \cdot 17014)}{19790^2} = 0.7780 \rightarrow$ the expected level of agreement due to chance is 77.8%

$\kappa = \frac{p_{\text{o}} - p_{\text{e}}}{1 - p_{\text{e}}} = \frac{0.8943 - 0.7780}{1 - 0.7780} = 0.524 \rightarrow$ according to the benchmark scale in Table 10.4 this represents a moderate level of agreement.

The kappa paradox

The κ has an important limitation: It is affected by the prevalence of the condition being studied. This causes an effect by which two raters can have high agreement but still emerge with a low kappa value. This problem is referred to as **the kappa paradox** (Feinstein and Cicchetti, 1990).

Illustrative Example 10.2 Kappa paradox

The data in Table 10.6 demonstrate a kappa paradox. Table 10.6A demonstrates observed proportions of agreement (p_{obs}) of 0.85 and kappa of 0.70 (''substantial agreement''). Table 10.6B also demonstrates observed proportions of agreement p_{obs} of 0.85, but in this case the kappa statistic is 0.32 (''fair agreement'').

Table 10.6 Demonstration of the kappa paradox (Feinstein and Cicchetti, 1990).

Table 10.6A

	Rater B		
Rater A	+	–	Total
+	40	9	49
–	6	45	51
Total	46	54	100

$p_{\text{obs}} = 0.85$
$p_{\text{exp}} = 0.50$
$\kappa = 0.70$

Table 10.6B

	Rater B		
Rater A	+	–	Total
+	80	10	90
–	5	5	10
Total	85	15	100

$p_{\text{obs}} = 0.85$
$p_{\text{exp}} = 0.78$
$\kappa = 0.32$

Several options have been offered as solutions to the kappa paradox. One approach uses alternative measures of agreement that are resistant to the kappa paradox. Two such alternatives are the **Brennan–Prediger kappa coefficient** (also called the **G index;** Holley and Guilford, 1964; Brennan and Prediger, 1981) and **Gwet's AC1** (Gwet, 2010).[a] These statistics can be calculated with WinPEPI's PairsEtc → ''A. 'Yes-no' (dichotomous) variable'' program.

One practical solution to the kappa paradox is to accompany the kappa statistic with the **proportion of specific positive agreement** (p_{pos}), which is

$$p_{pos} = \frac{2a}{f_1 + g_1} \tag{10.4}$$

and the **proportion of specific negative agreement** (p_{neg})

$$p_{neg} = \frac{2d}{f_2 + g_2} \tag{10.5}$$

Use of these statistics to complement κ provides a more complete picture of the agreement between the two raters.

Illustrative Example 10.3 Proportion of positive agreement and proportion of negative agreement

In Illustrative Example 10.1 (Table 10.3) we calculated and observed a level of agreement of 89.4% and κ statistic of 0.52. The proportion of positive agreement for these data is

$$p_{pos} = \frac{2a}{f_1 + g_1} = \frac{2 \cdot 1467}{2249 + 2776} = 0.584$$

The proportion of negative agreement is

$$p_{neg} = \frac{2d}{f_2 + g_2} = \frac{2 \cdot 16232}{17541 + 17014} = 0.939$$

This indicates that the agreement for positive diagnoses is inferior to that of negative diagnoses, suggesting that further work is needed to decrease the observers' disparities in the positive direction.

10.3 Validity

We use the term **validity** to describe the ability of a test or diagnostic procedure to accurately discriminate between people who do and do not have the disease of interest. A perfectly reliable and valid test would correctly discriminate between people with and without disease without fail.

We will discuss four measures of diagnostic test validity: sensitivity, specificity, predictive value positive, and predictive value negative. To calculate these measures, we must first classify test results into one of the following four categories:

[a] For an exhaustive consideration of kappa alternatives see Gwet, K.L. (2010). *Handbook of Inter-Rater Reliability*, 2nd edn, Advanced Analytics, LLC, Gaithersburg, MD.

- **True positives** (TP) have the disease in question and show positive test results.
- **True negatives** (TN) do not have the disease in question and show negative test results.
- **False positives** (FP) do not have the disease in question but show positive test results.
- **False negatives** (FN) have the disease but show negative test results.

This assumes there is a definitive ''gold standard'' means of identifying individuals who have and do not have the disease in question by which to make these classifications. After each result is classified into one of the above four categories, the frequency of results is cross-tabulated to form a table similar to the one shown in Table 10.7.

Sensitivity and specificity

Sensitivity (SEN) is the probability that a test result will be positive when the test is administered to people who actually have the disease or condition in question. Using conditional probability notation, we define sensitivity as $\Pr(T+ \mid D+)$, where Pr denotes ''probability,'' $T+$ denotes ''test positive,'' $D+$ denotes ''disease positive,'' and the vertical line (|) denotes ''conditional upon.'' Thereby, Pr ($T+ \mid D+$) is read as ''the probability of being test positive conditional upon being disease positive.''

Sensitivity is calculated by administering the test to subjects who have the disease in question. The number of diseased people who test positive is divided by the total number of diseased people tested:

$$\text{SEN} = \frac{\text{TP}}{\text{all those with the disease}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10.6}$$

Specificity (SPEC) is the probability that a test will be negative when administered to people who are free of the disease or condition in question. In other words, specificity is the probability of being test negative conditional upon being disease negative: $\text{SPEC} = \Pr(T- \mid D-)$.

Specificity is calculated by administering the test to disease-free subjects. The number of people testing negative is divided by the total number of disease-free people tested:

$$\text{SPEC} = \frac{\text{TN}}{\text{all those without the disease}} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{10.7}$$

Table 10.7 Notation for calculating sensitivity, specificity, predictive value positive, and predictive value negative.

	Disease +	Disease –	Total
Test +	TP	FP	TP + FP
Test –	FN	TN	FN + TN
	TP + FN	FP + TN	*N*

Illustrative Example 10.4 Teen smoking questionnaire (sensitivity and specificity)

To illustrate sensitivity and specificity, let us consider a hypothetical survey of teen smoking in which a questionnaire is used as a screening instrument to help determine whether subjects smoke. We are concerned that many teen smokers will feel compelled to falsely answer in the negative, so we compare the results of the questionnaire to a more reliable method of ascertainment based on testing for cotinine in the saliva. (Cotinine, a major detoxication product of nicotine, is a biomarker for tobacco smoke.) Thus, the questionnaire serves as a rapid and inexpensive screening toll and the salivary cotinine test serves as the ''gold standard'' method of ascertainment. Results of our study are shown in Table 10.8. Thus

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{65}{65 + 35} = 0.65$$

and

$$\text{SPEC} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{99}{99 + 1} = 0.99$$

Predictive value positive and predictive value negative

Although sensitivity and specificity quantify a test's accuracy in the presence of known disease status, they are unable to predict the performance of the test in the population. To accomplish this objective, the alternative indices of predictive value positive and predictive value negative are needed.

The **predictive value positive of a positive test** (PVPT) is the probability that a person with a positive test will actually have the disease in question. In other words, the predictive value positive is the probability of being disease positive conditional upon being test positive: PVPT = Pr($D+ \mid T+$). This statistic is calculated by dividing the number of true positives by all those people who test positive:

$$\text{PVPT} = \frac{\text{TP}}{\text{all those who test positive}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10.8}$$

The **predictive value of a negative test** (PVNT) is the probability that a person who shows a negative test will be disease negative—the probability of disease negative ''given'' test negativity: PVNT = Pr($D- \mid T-$). The predictive value negative is calculated by dividing the number of true negatives by all those people who test negative:

$$\text{PVNT} = \frac{\text{TN}}{\text{all those who test negative}} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{10.9}$$

Table 10.8 Data for Illustrative Examples 10.4–10.6. Results of a smoking survey questionnaire and definitive salivary cotinine test: fictitious data.

	Salivary cotinine test (''gold standard'')		
	+	–	Total
Response to questionnaire +	65	1	66
Response to questionnaire –	35	99	134
	100	100	200

The distinction between sensitivity/specificity and predictive value positive/predictive value negative may at first appear confusing. This becomes less confusing if one remembers that sensitivity and specificity quantify a test's accuracy given the known disease status of study subjects, whereas predictive values quantify a test's accuracy given only the test results.

Illustrative Example 10.5 Teen smoking questionnaire (PVPT and PVNT)

Let us return to the data in Illustrative Example 10.4 on the validity of data derived from a smoking questionnaire. Data are in Table 10.8. In this example, we have 65 true positives and 1 false positive. Therefore,

$$\text{PVPT} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{65}{65 + 1} = 0.985$$

This means that 98.5% of the study subjects that responded in the affirmative were actually smokers. The **false positive rate** is the complement of the PVPT. Therefore, the false positive rate was 1 – 0.985 = 0.015.

The questionnaire identified 35 false negatives and 99 true negatives. Since 99 of the 134 people who responded to the questionnaire in the negative were actual nonsmokers,

$$\text{PVNT} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{99}{99 + 35} = 0.739$$

This means that 73.9% of the negative responders were nonsmokers. The **false negative rate** is the complement of the PVNT. Therefore, the false negative rate was 1 – 0.739 = 0.261.

True prevalence and apparent prevalence

The prevalence of disease can be calculated on the basis of the true number of people with the disease in the population or the apparent number of people with the disease based on screening test results. The **true prevalence** of the disease (P) represents the proportion of people who actually have the disease or condition:

$$P = \frac{\text{all those with the disease}}{\text{all those tested}} = \frac{\text{TP} + \text{FN}}{N} \tag{10.10}$$

where TP represents the number of true positives, FN represents the number of false negatives, and N represents all those tested.

The **apparent prevalence** of a disease (P^*) represents the proportion of people who test positive on a screening test:

$$P^* = \frac{\text{all those who test positive}}{\text{all those tested}} = \frac{\text{TP} + \text{FP}}{N} \tag{10.11}$$

where TP represents the number of true positives, FP represents the number of false positives, and N represents all those tested.

Illustrative Example 10.6 Teen smoking questionnaire (true prevalence and apparent prevalence)

The apparent prevalence and true prevalence will differ when the screening test is imperfect. In Illustrative Examples 10.3 and 10.4 (Table 10.8), the true prevalence of smoking is

$$P = \frac{\text{TP} + \text{FN}}{N} = \frac{65 + 35}{200} = 0.500$$

In contrast, the apparent prevalence is

$$P^* = \frac{\text{TP} + \text{FP}}{N} = \frac{65 + 1}{200} = 0.333$$

This discrepancy is due to the under-reporting of smoking on the questionnaire.

Relation between prevalence and the predictive value of a positive test

The predictive value of a positive (PVPT) test depends on the sensitivity of the test, the specificity of the test, and the prevalence of the disease in the population in which the test is used. Although the first two determinants of predictive value (sensitivity and specificity) are not surprising, many students are caught off guard by the important role prevalence plays in determining predictive value. In general, if the prevalence of disease is low, the predictive value positive will be low. If the prevalence of disease is high, the predictive value positive will be high. This relationship holds for all diagnostic tests that fall short of perfection.

Illustrative Example 10.7 How the same test used in different populations can have quite different predictive values

Consider using a screening test with a sensitivity of 0.99 and specificity of 0.99 in two different populations. *Population A* has a prevalence of 1 in 10 (0.10). *Population B* has a prevalence of 1 in 1000 (0.001). Each population consists of 1 000 000 people.

Note that the number of people with disease in each population is equal to the prevalence of disease times the population size:

$$\text{No. with disease} = P \times N \tag{10.12}$$

Thus, *Population A* has 0.1 × 1 000 000 = 100 000 cases, and *Population B* has 0.001 × 1 000 000 = 1000 cases.

Because the SEN of the test is 99%, it correctly identifies 99 000 (99%) of the 100 000 cases in *Population A*. This leaves 1000 false negatives in this population. In addition, because the SPEC of the test is 99%, it correctly identifies 891 000 (99%) of the 900 000 non-cases as true negatives, leaving 9000 false positives. Table 10.9A shows the results of the test in *Population A* Using these results, PVPT in *Population A* is 91.7% (calculations below Table 10.9A).

Using the same type of reasoning, the test correctly identifies 990 (99%) of the 1000 cases and leaves 10 false negatives in *Population B*. It also correctly identifies 989 010 (99%) of the non-cases as true negatives in *Population B*, leaving 9900 false positives. The predictive value positive of the test in *Population B*, therefore, is only 9.0% (Table 10.9B). Thus, the PVPT is substantially lower in *Population B* than in *Population A*. This is because of *Population B*'s lower prevalence of disease.

Bayesian formulas for predictive value

The PVPT can be calculated directly from its sensitivity, specificity, and the prevalence of disease in the population it is being used in, if these quantities are known, according to the formula:

$$\text{PVPT} = \frac{(P)(\text{SEN})}{(P)(\text{SEN}) + (1 - \text{SPEC})(1 - P)} \tag{10.13}$$

Table 10.9 Data for Illustrative Example 10.7. Results of a screening test which has SEN = 0.99 and SPEC = 0.99 in two different populations.

Table 10.9A *Population A* (prevalence = 0.10)

	Disease +	Disease –	Total
Test +	99 000	9000	108 000
Test –	1000	891 000	892 000
	100 000	900 000	1 000 000

$$\text{PVPT} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{99\,000}{99\,000 + 9000} = 0.917$$

Table 10.9B *Population B* (prevalence = 0.001)

	Disease +	Disease –	Total
Test +	990	9990	10 980
Test –	10	989 010	989 020
	1000	999 000	1 000 000

$$\text{PVPT} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{990}{990 + 9900} = 0.090$$

where PVP represents predictive value positive, P represents (true) prevalence, SEN represents sensitivity, and SPEC represents specificity. Because Formula (10.13) is derived using Bayes's law of probability, it is called the ''Bayesian formula for predictive value positive.''

Illustrative Example 10.8 Teen smoking questionnaire (PVPT with the Bayesian formula)

Formula (10.13) is used to calculate the PVPT of the data in Table 10.6. Given the test's sensitivity of 0.650 and specificity of 0.990, and the population prevalence of 0.500,

$$\text{PVPT} = \frac{(P)(\text{SEN})}{(P)\,(\text{SEN}) + (1 - \text{SPEC})(1 - P)} = \frac{(0.500)\,(0.650)}{(0.500)\,(0.650) + (1 - 0.990)\,(1 - 0.500)} = 0.985$$

This calculated value matches the previously calculated value determined in Illustrative Example 10.5.

The Bayesian formula for the PVPT allows us to plot the predictive value of a positive test as a function of prevalence, sensitivity, and specificity. Figure 10.3 plots this relation for three different diagnostic tests. The sensitivity of all three tests is held constant at 0.99. Specificity varies between 0.80 and 0.99, as labeled in the figure. This figure indicates that all three tests have low predictive value positive when used in populations with low disease prevalence and that the predictive value positive increases as a function of prevalence. It also indicates that tests of low specificity add little new information about the population when the prevalence of disease is low.

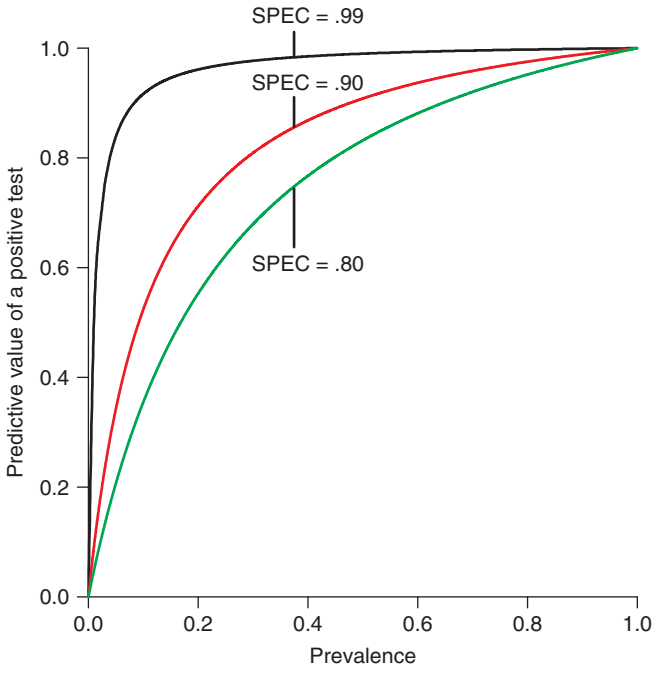


Figure 10.3 PVPT as a function of prevalence. All three tests have a sensitivity of 0.99. Tests of three specificities are considered (indicated by SPEC).

Relation between prevalence and the predictive value of a negative test

As was the case with the predictive value of a positive test, the predictive value negative of a negative test also depends on the sensitivity and specificity of the test and prevalence of disease in the population in which the test is being used. The Bayesian formula relating these factors is

$$\text{PVNT} = \frac{(1-P)(\text{SPEC})}{(1-P)\,(\text{SPEC}) + (1-\text{SEN})(P)} \tag{10.14}$$

where PVN represents predictive value negative, P represents (true) prevalence, SEN represents sensitivity, and SPEC represents specificity.

Application of this formula to the illustrative teen smoking data results in

$$\text{PVNT} = \frac{(1-P)(\text{SPEC})}{(1-P)\,(\text{SPEC}) + (1-\text{SEN})(P)} = \frac{(1-0.500)\,(0.990)}{(1-0.500)\,(0.990) + (1-0.650)\,(0.500)}$$
$$= 0.739$$

Figure 10.4 plots the relation between prevalence and the predictive value of a negative test for three different diagnostic tests. Each test has a specificity of 0.99. The sensitivity of the test varies from 0.80 to 0.99, as indicated in the figure. This figure shows how PVNT decreases as a function of prevalence. There is also a correlation

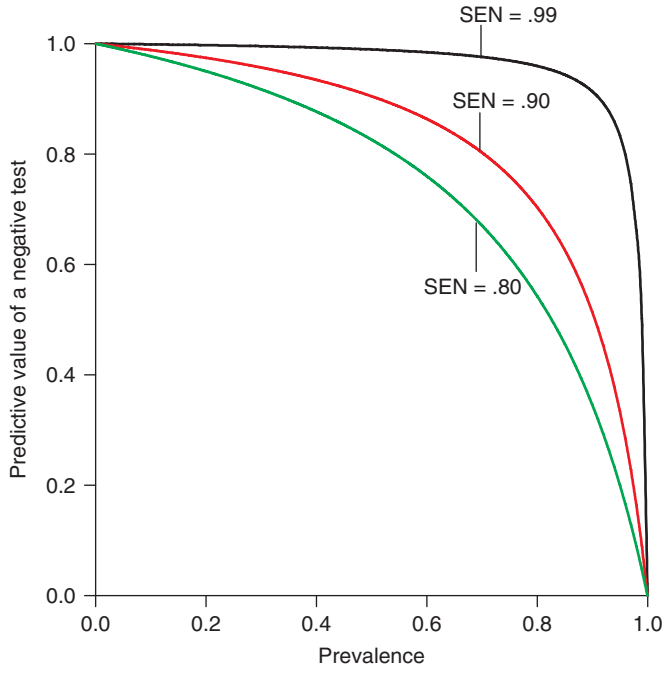


Figure 10.4 PVNT as a function of prevalence. All three tests have a specificity of 0.99. Tests of three sensitivities are shown.

between sensitivity and predictive value negative: low sensitivity is associated with low PVNT.

Selecting a cutoff point for positive and negative test results

Many diagnostic test results are based on a continuum of values that must be converted into either ''positive'' or ''negative'' results in order to be interpreted. For example, enzyme-linked immunosorbent assay methods used to detect the presence of human immunodeficiency virus (HIV) indicate the presence of the HIV antibody by glowing or showing a color (''fluorescing''). Higher concentrations of HIV antibody are associated with greater levels of fluorescence. The degree of fluorescence is read on a numerical optical density ratio (ODR) scale that demonstrates a range. Owing to nonspecific immunologic reactions, sera from people free from HIV infection may also demonstrate some degree of nonspecific immunofluorescence upon testing. It would be convenient if the distributions of ODR values for HIV-positive and HIV-negative populations did not overlap. Unfortunately, this is not the case. As a result, sera from some HIV-negative people demonstrate higher ODR readings than their HIV-positive counterparts (Figure 10.5).

At what point do we declare a fluorescing HIV antibody test ''positive''? Selecting a low cutoff point would identify all people carrying HIV as positive (Figure 10.5*a*). However, in so doing, false positives will also be identified. The resulting test will

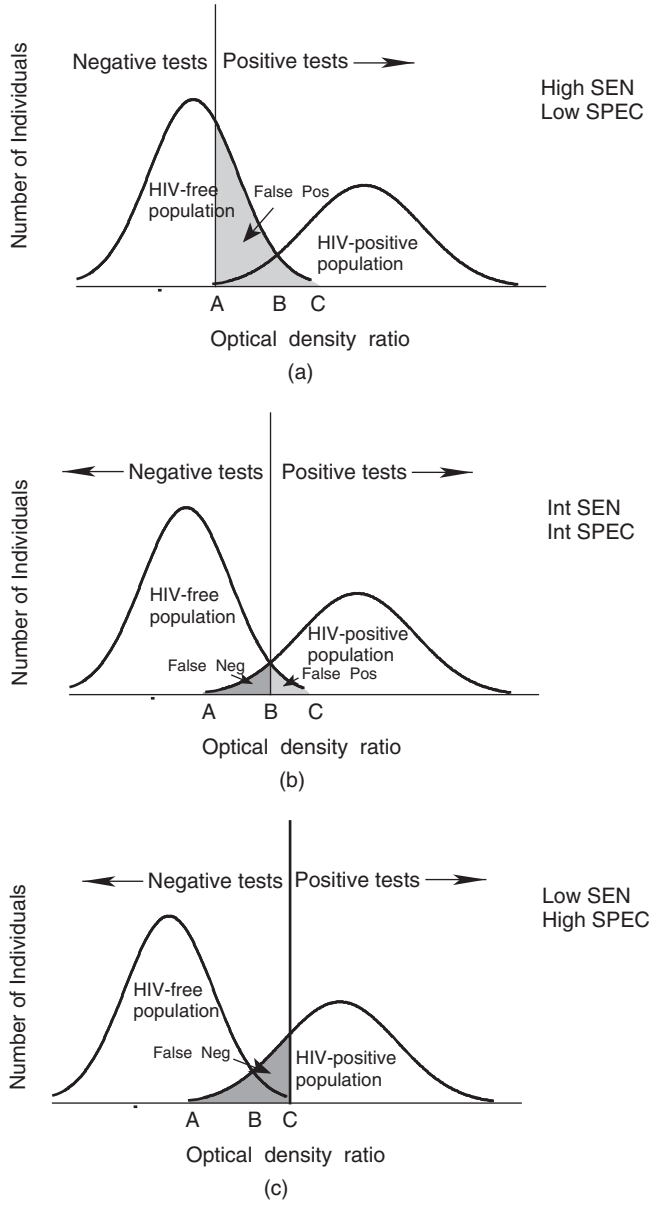


Figure 10.5 Effect of setting different cutoffs on the sensitivity and specificity of an enzyme-linked immunosorbent HIV assay—hypothetical data.

have high sensitivity but compromised specificity. Selecting an intermediate cutoff results in a test that identifies fewer false positives, but some false negatives will now be identified. The resulting test will have intermediate sensitivity and intermediate specificity (Figure 10.5*b*). Selecting a high cutoff point eliminates false positives; however, now many more false negatives will be identified. The resulting test will

have high specificity but low sensitivity (Figure 10.5*c*). Thus, the selection of the cutoff point determines the sensitivity and specificity of the test.

The selection of the cutoff point should reflect the purpose of the test. When attempting to avoid false negatives, a low cutoff point is used (Figure 10.5*a*). When attempting to minimize overall error, an intermediate cutoff point is used (Figure 10.5*b*). When attempting to avoid false positives, a high cutoff point is used (Figure 10.5*c*). Each strategy has its place as well as its consequences.

Given the inherent trade-offs in sensitivity and specificity—tests that are highly sensitive tend to be less specific than tests that are highly specific, and vice versa—a multistage screening program is often used to accurately identify cases. In its simplest form, a two-stage process is used. If we hope to avoid false negatives, the first stage of screening uses a highly sensitive test. In using a highly sensitive test, false negatives are avoided, but false positives may be commonplace. Thus, a second stage of screening is needed. This second stage sorts out true positives and false positives using a test of high specificity. This process is analogous to casting a wide net and sorting out the true positives from false positives later (Figure 10.6). Because the

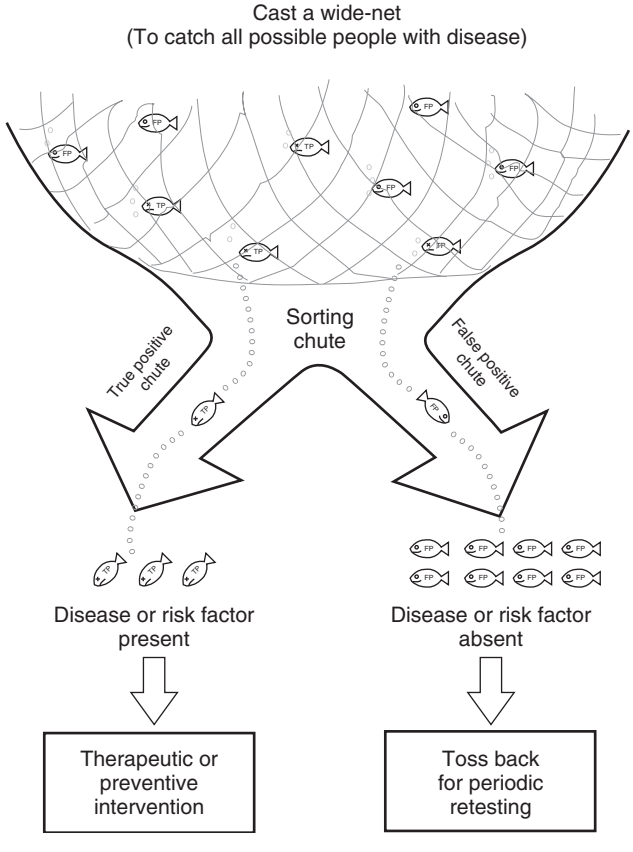


Figure 10.6 Fishing metaphor for two-stage screening. First stage of screening casts a wide net to identify all possible cases (i.e., uses a test of high sensitivity). Second stage uses a test of high specificity to sort out true positives from false positives.

purpose of screening is to identify cases for early diagnosis and treatment, non-cases are periodically re-examined to determine whether disease has developed since the last screening procedure.

Summary

Key points

1 The accuracy of a diagnostic procedure is affected by two factors: (a) the agreement of results on replication (reliability) and (b) the ability to discriminate between people with and without disease (validity). The investigator must take all practical steps to increase the reproducibility and validity of results by standardizing the conditions under which data are collected. Furthermore, (s)he must take all necessary steps to quantify and understand deviations from the true biological state of affairs, whenever possible.
2 The kappa statistic (κ) is a chance-corrected measure of reliability. Guides for interpreting kappa are listed in Table 10.4. However, because of the kappa paradox, the interpretation of kappa requires caution and supplementary measures of agreement in the form of the proportion of positive agreement (p_{pos}) and proportion of negative agreement (p_{neg}).
3 SEN, SPEC, PVPT, and PVNT are measures of diagnostic test validity. Using conditional probability notation, these terms are defined as follows:
   - sensitivity = $\Pr(T+ | D+)$
   - specificity = $\Pr(T- | D-)$
   - predictive value positive = $\Pr(D+ | T+)$
   - predictive value negative = $\Pr(D- | T-)$.
4 The PVPT is a function of a test's sensitivity, specificity, and the prevalence of the disease in the population in which it is used. PVPTs will tend to be low in populations with a low prevalences of disease.
5 In selecting a cutoff point for determining a positive test result, there are trade-offs in opting for either a sensitive or specific test. To circumvent this problem, screening programs use several stages of screening, initially using a sensitive test to identify all possible cases, followed by more specific procedures to sort out true positives and false positives.

Reliability notation

a	Number of subjects in which both raters offer a positive diagnosis
b	Number of subjects in which rater A offers a positive diagnosis and rater B offers a negative diagnosis
c	Number of subjects in which rater B offers a positive diagnosis and rater A offers a negative diagnosis
d	Number of subjects in which both raters offer negative diagnoses
f_1	Number of rater B's diagnoses that are positive

f_2	Number of rater B's diagnoses that are negative
g_1	Number of rater A's diagnoses that are positive
g_2	Number of rater A's diagnoses that are negative
p_{obs}	Observed proportion of agreement (Formula 10.1)
p_e	Expected proportion of agreement due to chance (Formula 10.2)
κ	Kappa statistic (Formula 10.3)
p_{pos}	Proportion of specific positive agreement (Formula 10.4)
p_{neg}	Proportion of specific negative agreement (Formula 10.5)

Validity notation

TP	True positive
TN	True negative
FP	False positive
FN	False negative
N	Sample size
SEN	Sensitivity (Formula 10.6)
SPEC	Specificity (Formula 10.7)
PVPT	Predictive value of a positive test (Formula 10.8; Formula 10.13 (Baysian))
PVNT	Predictive value of a negative test (Formula 10.9; Formula 10.14 (Baysian))
P	True prevalence (Formula 10.10)
P^*	Apparent prevalence (Formula 10.11)

Exercises

10.1 Sign, symptom, or test? Determine whether each of the following is primarily a symptom, sign, or test.

(A) Chills
(B) Fever of 104.6 °F
(C) Sore throat (sensation)
(D) Visibly swollen and reddened lining of the throat.

10.2 Cross-tabulate first. We wish to compare the results of a new screening test (Test B) with the current gold standard method (Test A). Data appear in Table 10.10.

(A) How many false positives did screening test B ascertain?
(B) How many false negatives were evident?
(C) Cross-tabulate the data to form a two-by-two table similar to that of Table 10.7.
(D) What is the sensitivity of test B?
(E) What is the specificity of test B?

10.3 Sensitivity, specificity, and predictive value. To characterize the sensitivity and specificity of a simple and inexpensive screening test, 200 people

Table 10.10 Data for Exercise 10.2. Results from a gold standard procedure (test A) and a new screening test (test B).

Specimen number	Test A (''gold standard'')	Test B (screening test)
1	–	+
2	+	+
3	–	–
4	+	+
5	–	–
6	+	+
7	–	–
8	–	–
9	–	+
10	+	–

Table 10.11 Data for Exercise 10.3.

	Definitive exam		
	+	–	
Screening test +	40	25	65
Screening test –	10	125	135
	50	150	200

participate in a study in which they receive the screening test while simultaneously undergoing a definitive clinical examination. Results are tabulated in Table 10.11.

(A) What is the sensitivity of the screening test?
(B) What is the specificity of the screening test?
(C) What is the PVP of the screening test?
(D) What is the PVN of the screening test?

10.4 Bayesian formula. A screening program aimed at the early detection of a cancer uses a screening test that demonstrates a sensitivity of 0.95 and specificity of 0.90. Of those attending the screening program, 1 per 1000 (0.001) actually has the cancer in question. What proportion of those who screen positively have the cancer in question? Use Bayesian Formula (10.13) to derive the numerical answer. Provide a brief interpretation of the results.

10.5 Drug test. Suppose 5% of the people in a study population use illicit drugs. We employ a test that is 95% ''accurate'' in that the test will be positive in 95% of drug users and negative in 95% of nonusers. If a person selected at random from the population demonstrates a positive test result, what is the likelihood that the person is actually a drug user?

10.6 Two raters. Results from two independent raters are shown in Table 10.12. How well do the raters agree?

Table 10.12 Data for Exercise 10.6.

	Rater B		
Rater A	+	–	Total
+	150	31	181
–	28	239	267
Total	178	270	448

Table 10.13 Data for Exercise 10.7.

	Definitive diagnosis		
Test	+	–	Total
+	15	7	22
–	3	145	148
Total	18	152	170

Table 10.14 Data for Exercise 10.8.

	Tornado occurrence		
Prediction	+	–	Total
+	11	14	25
–	3	906	909
Total	14	920	934

Comment: Try using *www.OpenEpi.com* → Counts → Screening Calculator'' to check your answer.

10.7 Sensitivity and specificity. The results of a screening test are compared with definitive diagnoses in 170 patients with results shown in Table 10.13.

(A) Calculate this test's sensitivity and specificity.

(B) Why is Exercise 10.6 a reproducibility analysis, while this analysis is a validity analysis?

10.8 Predicting tornadoes. In July of 1884, J.P. Finley of the U.S. Army Signal Corps published a paper in which he summarized results of a forecasting program intended to predict tornados (Murphy, 1996). Finley's data for April 1883 are shown in Table 10.14.

(A) From this table we can see that Finley's predictions were correct in 917 of the 934 (98.2%) instances. However, quite a few of these predication could have been lucky guesses.[b] Calculate κ to determine the extent to which Finley's predictions exceed that predicted by chance.

(B) How sensitive were Finley's predictions? Interpret this result.

(C) Calculate the PVPT and PVNT of Finley's predictions and comment on these findings.

[b] Consistently predicting ''no tornado'' would derive $p_0 = 920/934 = 98.5\%$.

Table 10.15 Data for Exercise 10.9.

	Pathologist B		
Pathologist A	+	–	Total
+	9	2	11
–	0	489	489
Total	9	491	500

10.9 Screening for bladder cancer. A screening test for bladder cancer uses the staining properties of exfoliated cells in the urine to detect potential bladder cancer cases. Two pathologists review 500 samples and come up with the diagnostic results shown in Table 10.15. (Data are fictitious.)

(A) Explain why this is a reproducibility analysis and not a validity analysis.

(B) Assess the reproducibility of the results.

(C) Further research on the accuracy of the screening test demonstrates that it has a sensitivity of 90% and specificity of 98%. Suppose this test is used in 100 000 individuals from a population in which the prevalence of subclinical bladder cancer is 0.1%. Set up a two-way table showing the number of true positives, true negatives, false positives, and false negatives expected when using this screening test in this population.

(D) What is the predictive value of a positive test in this population?

(E) What is the predictive value of a negative test in this population?

(F) The same test is used in a patient population demonstrating chronic hematuria and other symptoms of possible bladder cancer. The prevalence of bladder cancer in this clinical population is 1 in 10. Set up a two-way table showing the distribution of true positives, true negatives, false positives, and false negatives expected when using the test in 1000 people from this clinical population.

(G) What is the predictive value of a positive test when used in the clinical population described in item (F)?

(H) Why is the predictive value positive of the test in the clinical population so much greater than in the general population?

10.10 Leg length inequality. A study evaluated the inter-examiner reliability of a leg length determination in which the longer leg was ascertained by experienced and inexperienced chiropractors (Holt *et al.*, 2009). The inexperienced chiropractors had undergone extensive training before the study. Examinations

Table 10.16 Data for Exercise 10.10.

	Rater B		
Rater A	+	–	Total
+	12	2	14
–	4	28	32
Total	16	30	46

were carried out in the prone position. Data in the table below represent agreements between the raters. The + sign indicates that the left leg was deemed to be shorter; the − sign indicates that the right leg was deemed shorter. Evaluate the reproducibility of these results using the data in Table 10.16.

Review questions

R.10.1 Distinguish between reliability and validity.

R.10.2 Distinguish between a sign and a symptom.

R.10.3 What statistics are used to quantify the reliability of a diagnostic measure? What statistics are used to quantify validity?

R.10.4 What is the advantage of using the kappa statistic instead of the observed percentage of agreement when assessing reliability?

R.10.5 True or false? A kappa statistic of 0 indicates no agreement. Explain your response.

R.10.6 Define: (a) true positive, (b) true negative, (c) false positive, and (d) false negative.

R.10.7 Match the statistic with the conditional probability it estimates. *Statistic*: SEN, SPEC, PVPT, PVNT *Conditional probability*: (a) Pr(T−|D−), (b) Pr(T+|D+), (c) Pr(D+|T+), and (d) Pr(D−|T−)

R.10.8 What is the term is used to refer to the expected proportion of disease-free individuals that will have a negative test?

R.10.9 What is the term is used to refer to the expected proportion of negative tests that are disease-free?

R.10.10 What factors determine PVPT?

R.10.11 Explain why the PVPT tends to be low when prevalence is low.

R.10.12 Consider a test in which low values indicate the presence of a condition. Assume there is overlap in this test's results in the diseased and non-diseased populations. What effect will decreasing the cut-off point for a positive test have on the sensitivity and specificity of the test?

R.10.13 Assuming the results of a positive test will ultimately result in no negative effects and minimal costs, should the first stage of a screening program be highly sensitive or highly specific? Explain.

References

Ahlbom, A., and Norell, S. (1990) *Introduction to Modern Epidemiology*, 2nd edn, Epidemiology Resources, Chestnut Hill, MA.

Björkhem, I., Bergman, A., Falk, O., Kallner, A., Lantto, O., Svensson, L., Äkerlöf, E., and Blomstrand, R. (1981) Accuracy of some routine methods used in clinical chemistry as judged by isotope dilution—mass spectrometry. *Clinical Chemistry*, **27**, 733–735.

Brennan, R.L., and Prediger, D.J. (1981) Coefficient kappa, some uses and alternatives. *Educational and Psychological Measurement*, **41** (3), 687–699.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20** (1), 37–46.

Commission on Chronic Illness. (1957) *Chronic Illness in the United States Vol. IV, Chronic Illness in a Large City: The Baltimore Study*, Harvard University Press, Cambridge, MA.

Feinstein, A.R., and Cicchetti, D.V. (1990) High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, **43** (6), 543–549.

Finley, J.P. (1884) Tornado predictions. *American Meteorological Journal*, **1**, 85–88.

Gwet, K.L. (2010) *Handbook of Inter-Rater Reliability*, 2nd edn, Advanced Analytics, LLC, Gaithersburg, MD.

Holley, J.W., and Guilford, J.P. (1964) A note on the G index of agreement. *Educational and Psychological Measurement*, **24**, 749–753.

Holt, K.R., Russell, D.G., Hoffmann, N.J., Bruce, B.I., Bushell, P.M., and Taylor, H.H. (2009) Interexaminer reliability of a leg length analysis procedure among novice and experienced practitioners. *Journal of Manipulative Physiological Therapeutics*, **32** (3), 216–222.

Landis, J.R., and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33** (1), 159–174.

Lilienfeld, A.M., and Kordan, B. (1966) A study in the variability in the interpretation of chest X-rays in the detection of lung cancer. *Cancer Research*, **26**, 2145–2147.

Lilienfeld, A.M., and Lilienfeld, D.E. (1980) *Foundations of Epidemiology*. 2nd edn, Oxford University Press, New York.

Murphy, A.H. (1996) The Finley Affair: A Signal Event in the History of Forecast Verification. *AMS: Forecasting (March 96)*, **11** (1), 3–20.

Tukey, J.W. (1969) Analyzing data: Sanctification or detective work? *American Psychologist*, **24**, 83–91.

10.4 Chapter addendum (case study)

Screening for antibodies to the human immunodeficiency virus

Source: Centers for Disease Control. *1992 EIS Course*.
Authors: L. Peterson, G. Birkhead, and R. Dicker

Objectives

After completing this case study, the student should be able to:

1 Define and perform calculations of sensitivity, specificity, predictive value positive, and predictive value negative.
2 Describe the relationship between prevalence and predictive value.
3 Discuss the trade-offs between sensitivity and specificity.
4 List the principles of a good screening program.

Part I

In December 1982, a report in the *Morbidity and Mortality Weekly Report* (MMWR) described three persons who had developed acquired immunodeficiency syndrome (AIDS) but who had neither of the previously known risk factors for the disease: homosexual/bisexual activity with numerous partners and intravenous drug use. These three persons had previously received whole-blood transfusions. By 1983, widespread recognition of the problem of transfusion-related AIDS led to controversial recommendations that persons in known high-risk groups voluntarily defer from donating blood. In June 1984, after the discovery of the human immunodeficiency virus (HIV), five companies were licensed to produce enzyme-linked immunosorbent assay (EIA, then called ELISA) test kits for detecting the HIV antibody. A Food and

Drug Administration (FDA) spokesperson stated that ''getting this test out to the blood banks is our No. 1 priority.'' Blood bank directors were anxiously waiting to start screening blood with the new test, until 2 March 1985, the date the first test kit was approved by the FDA.

In the prelicensure evaluation, sensitivity and specificity of the test kits were estimated using blood samples from four groups: those with AIDS by the Centers for Disease Control and Prevention (CDC) criteria, those with other symptoms and signs of HIV infection, those with various autoimmune disorders and neoplastic diseases that could give a false-positive test result, and presumably healthy blood and plasma donors.

Numerous complex issues were discussed even before licensure. Among them were agreeing on the significance of a negative blood test, determining the percentage of antibody-positive persons who were capable of transmitting the virus, understanding the magnitude of the problem of false-positive test results, and determining whether test-positive blood donors should be notified.

It is now 2 March 1985, and you are the state epidemiologist of State Y. The first HIV antibody test kits will arrive in blood banks in your state in a few hours. Meeting with you to discuss the appropriate use of this test are the commissioner of health, the medical director of the regional blood bank, and the chief of the State Y Drug Abuse Commission.

To help in your discussions, you turn to prelicensure information regarding the sensitivity and specificity of test kit A. The information indicates that the sensitivity of test kit A is 95.0% (0.95) and the specificity is 98.0% (0.98).

Question 1: With this information, by constructing a two-by-two table, calculate the predictive value positive and predictive value negative of the EIA in a hypothetical population of 1 000 000 blood donors. Using a separate two-by-two table, calculate PVP and PVN for a population of 1000 drug users. Assume that the actual prevalence of the HIV antibody among blood donors is 0.04% (0.0004) and that of intravenous drug users is 10.0% (0.10). The blood bank director wants your assistance in evaluating the EIA as a test for screening donor blood in State Y. In particular, she is concerned about the possibility that some antibody-positive units will be missed by the test, and she wonders about false-positive test results since she is under pressure to develop a notification procedure for EIA-positive donors.

Question 2: Do you think that the EIA is a good screening test for the blood bank? What would you recommend to the blood bank director about notification of EIA-positive blood donors? The chief of the State Y Drug Abuse Commission has noticed a dramatic increase in AIDS among clients of his intravenous-drug-abuse treatment programs. He wants to do a voluntary HIV antibody seroprevalence survey of intravenous-drug-abuse clients for planning purposes and would like to assess the feasibility of using the test as part of behavior modification counseling.

Question 3: Do you think that the EIA performs well enough to justify informing test-positive clients in the drug-abuse clinics that they are positive for HIV?

Question 4: If sensitivity and specificity remain constant, what is the relationship of prevalence to predictive value positive and predictive value negative?

Part II

EIA results are recorded as optical density ratios (ODRs). The ODR is a ratio of absorbance of the tested sample to the absorbance of a control sample. The greater the ODR, the more ''positive'' is the test result. The EIA, as with most other screening tests, is not perfect; there is some overlap of optical density ratios of samples that are actually antibody positive and those that are actually antibody negative. This is illustrated in Figure 10.5.

Establishing the cutoff value to define a positive test result from a negative one is somewhat arbitrary. You initially decide that optimal density ratios greater than B on Figure 10.5 are positive.

Question 5a: In terms of sensitivity and specificity, what happens if you raise the cutoff from B to C in Figure 10.5?

Question 5b: In terms of sensitivity and specificity, what happens if you lower the cutoff from B to A in Figure 10.5?

Question 6: From what you know now, what is the relationship between sensitivity and specificity of a screening test?

Question 7: What would the blood bank director and the head of drug treatment consider in deciding where the cutoff point should be for each program? Who would probably want a lower cutoff value?

Part III

You are concerned that because of the low predictive value of the EIA in the blood donor population, the blood bank personnel cannot properly inform those who are EIA positive of their actual antibody status. For this reason, you wish to evaluate the Western blot test as a confirmatory test for HIV antibody.

The Western blot test identifies antibodies to specific proteins associated with the HIV. The Western blot is the most widely used secondary test to detect HIV antibody because its specificity exceeds 99.99%; however, it is not used as a primary screening test because it is expensive and technically difficult to perform. Its sensitivity is thought to be lower than that of the EIA.

Because the Western blot test is not generally available, the blood bank director is wondering whether the initial EIA-positive results can be confirmed by repeating the EIA and by considering persons to have the antibody only if results of both tests are positive.

You decide that you want to compare the performance of the repeat EIA and the Western blot as confirmatory tests. To do this, use your earlier hypothetical sample of 1 000 000 blood donors. Assume that serum specimens that are initially positive by EIA are then split into two aliquots; a repeat EIA is performed on one portion and a Western blot on the other portion.

Question 8: What is the actual antibody prevalence in the population of persons whose blood samples will receive confirmatory testing?

Question 9: Calculate the predictive value positive of the two sequences of tests; EIA–EIA and EIA–Western blot (WB). Assume that the sensitivity and specificity of the EIA are 95.0 and 98.0%, respectively. Assume that the sensitivity and specificity of the Western blot are 80.0 and 99.99%, respectively.

Question 10: Why does the predictive value positive increase so dramatically with the addition of a second test? Why is the predictive value positive higher for the EIA–WB sequences than for the EIA–EIA sequence?

Part IV

It is now July 1987 and the governor has asked you to evaluate a proposed premarital HIV-antibody-screening program. A bill to establish the program is to be presented to the state legislature tomorrow. You estimate that 60 000 people will get married in your state in the next year. The proposed legislation requires that each prospective bride and groom submit a blood sample for EIA testing. Those positive EIA test results will then receive a Western blot test.

You decide that a goal of the screening program is to decrease inadvertent perinatal or sexual HIV transmission by determining who among those to be married are probably infected with the virus.

Question 11: What criteria would you consider in evaluating this proposed screening program? Tables 10.17 and 10.18 show the results of the testing, assuming that persons getting married have the same actual HIV antibody prevalence as blood donors (0.04%). In 1987 the sensitivity and specificity of the currently available version of EIA test kit A were 97.0 and 99.8%, respectively. The Western blot sensitivity and specificity were 95.0 and 99.99%, respectively.

With sequential testing, the sensitivity is 92%, the specificity is 100%, and predictive value positive is 100%.

Question 12: Compute the cost of the screening program. Assume a cost of $50.00 for every initial EIA test ($10.00 lab fee and $40.00 health-care-provider visit) and an additional $100.00 for EIA-positive persons who will need additional testing. What is the cost of the screening program in the next year? What is the cost per identified antibody-positive person?

Question 13: What is your final recommendation to the governor?

Table 10.17 Results of initial EIA test in people getting married.

	Actual antibody status		
	+	−	Total
+	23	120	143
Initial EIA	1	59 856	59 857
−	24	59 976	60 000

Table 10.18 Results of follow-up Western blot test in people getting married.

	Actual antibody status		
	+	−	Total
+	22	0	22
Follow-up Western blot	1	120	121
−	23	120	143

Additional information

The following ten principles of good mass screening programs were proposed by Wilson and Jungner (1968):

1. The condition being sought is an important health problem for the individual and the community.
2. There is an acceptable form of treatment for patients with recognizable disease.
3. The natural history of the condition, including its development from latent to declared disease, is adequately understood.
4. There is a recognizable latent or early symptomatic stage.
5. There is a suitable screening test or examination for detecting the disease at the latent or early symptomatic state, and this test is acceptable to the population.
6. The facilities required for diagnosis and treatment of patients revealed by the screening program are available.
7. There is an agreed policy on which to base treatment of patients.
8. Treatment at the presymptomatic, borderline stage of a disease favorably influences its course and prognosis.
9. The cost of the screening program (which would include the cost of diagnosis and treatment) is economically balanced in relation to possible expenditure on medical care as a whole.
10. Case finding is a continuing process, not a ''once and for all'' project.

Further reading—screening for HIV

Check, W.A. (1983) Preventing AIDS transmission: Would blood donors be screened? *JAMA*, **249**, 567–570.

Cleary, P.D., Barry, M.J., Mayer, K.H., *et al*. (1987) Compulsory premarital screening for the human immunodeficiency virus: Technical and public health considerations. *JAMA*, **258**, 1757–1762.

Goldsmith, M.F. (1985) HTLV-III testing of donor blood imminent; complex issues remain. *JAMA*, **253**, 173–181.

Marwick, C. (1985) Use of AIDS antibody test may provide more answers. *JAMA*, **253**, 1694–1699.

Schwartz, J.S., Dans, P.E., and Kinosian, B.P. (1988) Human immunodeficiency virus test evaluation, performance, and use: Proposals to make good tests better. *JAMA*, **259**, 2574–2579.

Sivak, S.L., and Wormser, G.P. (1986) Predictive value of a screening test for antibodies to HTLV-II. *American Journal of Clinical Pathology*, **85**, 700–703.

Sloand, E.M., Pitt, E., Chiarello, R.J., and Nemo, G.J. (1991) HIV testing: State of the art. *JAMA*, **266**, 2861–2866.

Turnock, B.J., and Kelly, C.J. (1989) Mandatory premarital testing for human immunodeficiency virus: The Illinois experience. *JAMA*, **261**, 3415–3418.

Ward, J.W., Grindon, A.J., Feorino, P.M., Schable, C., Parvin, M., and Allen, J.R. (1986) *JAMA*, **256**, 357–361.

Wilson, J.M.G., and Jungner, G. (1968) *Principle and practice of screening for disease*, World Health Organization, Geneva.

Further reading—general concepts of screening

Seilker, H.P. (1986) Clinical prediction rules. *New England Journal of Medicine*, **314**, 714.

Tyler, J.W., and Cullor, J.S. (1989) Titers, tests, and truisms: Rational interpretation of diagnostic serologic testing. *Journal of the American Veterinary Medical Association*, **194**, 1550–1558.

Table 10.19 Screening in blood bank population.

	Disease +	Disease −	Total
Test +	380	19 992	20 372
Test −	20	979 608	979 628
	400	999,600	1 000 000

Answers to case study: screening for antibodies to the human immunodeficiency virus

Part I

Answer 1

Blood bank calculations

Given: N = 1 000 000
EIA sensitivity 95.0%
EIA specificity 98.0%
Prevalence of HIV in blood donors is 0.04% (0.0004)
See Table 10.19 for numerical results.

Notes:

- The number who are antibody-positive is 1 000 000 × 0.0004 = 400.
- The number who are antibody-negative is 1 000 000 − 400 = 999 600.
- The number who are truly positive and who test positive is calculated as the left column total times sensitivity, or 400 × 0.95 = 380.
- The number of false negatives is calculated as 400 − 380 = 20.
- The number who are truly negative and who test negative is calculated as the right-column total times specificity, or 999 600 × 0.98 = 979 608.
- The number of false positives is calculated as 999 600−979 608 = 19 992.
- Row totals are next: 20 372 and 979 628.
- Now review formulas for PVP and PVN and calculate them.

$$\text{PVP} = 380/20\,372 = 0.019$$

$$\text{PVN} = 979\,608/979\,628 = 0.99998$$

Drug clinic calculations

Given: EIA sensitivity 95.0%
EIA specificity 98.0%
Prevalence of HIV in drug users is 10% (0.10)
See Table 10.20 for numerical results.

$$\text{PVP} = 95/113 = 0.841$$

$$\text{PVN} = 882/887 = 0.994$$

Answer 2 At the blood bank, the primary concern is the safety of the blood supply. The EIA is a good but not perfect screening for the blood bank. Ninety-five percent (380/400) of the antibody-positive units will be screened out, and 2%

Table 10.20 EIA screening in drug clinic population.

	Disease +	Disease –	Total
Test +	95	18	113
Test –	5	882	887
	100	900	1000

(20 372/1000 000) of the donated units will need to be discarded. Because only 1.9% of the test-positive persons will actually have the antibody (predictive value positive = 0.019), test-positive blood donors should *not* be notified on the basis of this test alone.

Answer 3 For the drug clinic clients, persons with a positive test will have an 84.1% chance of actually having the antibody (predictive value positive), while those with a negative test will only have a 0.6% chance of having the antibody (1 – PVN). Although the EIA is much more useful in separating those with and without antibody in the drug clinic than in the blood bank, 16% (1 – PVP) of drug clinic clients with a positive test result will not actually have the antibody (false-positive).

Note:

However, regardless of the test results, counseling of this population is important because they are engaging in high-risk behavior.

Answer 4 If the prevalence is high, the predictive value positive will be high, and the predictive value negative will be low. If the prevalence is low, the predictive value positive will be low, and the predictive value negative will be high (see Figures 10.3 and 10.4).

Part II

Answer 5a Moving the cutoff from B to C will decrease the sensitivity and will increase the specificity of the test.

Answer 5b Moving the cutoff from B to A will increase the sensitivity and will decrease the specificity of the test.

Answer 6 By changing the cutoff, if the sensitivity is increased, the specificity is decreased. Conversely, if the sensitivity is decreased, the specificity is increased.

Answer 7 The blood bank director's primary goal is to screen out antibody-positive (probably capable of transmitting the infection) blood at almost any cost. Therefore, (s)he would choose to have a very sensitive test. The cost will be a lower specificity; hence there will be more false-positive test results, and more blood will be discarded because of false-positive results. Because of the severe ramifications of notifying a person that he or she has the antibody, when, in fact, he or she does not (false-positive), the director of drug treatment will want a test with high specificity in order to maximize the predictive value positive.

For these reasons, the blood bank director will probably want a lower cutoff.

Part III

Answer 8 In this problem, all persons with a positive EIA result will receive Western blot confirmatory testing. From the hypothetical 1 000 000-person blood donor population in Question 1, 20 372 persons will have a positive test result. Of these 20 372 persons, 380 (1.9%) will actually have the antibody.

Answer 9 In this problem, we are assuming that both tests are independent—the results of the first test do not affect the results of the second test. This is generally not true with series of screening tests; the second test will not perform as well in a population that has already been screened with an initial test. Therefore, our calculations in this problem will overestimate the predictive value positive.

An example of nonindependence is the repeat EIA. On the initial EIA, some of the false-positive test results will be due to laboratory errors that will be unlikely to be repeated, such as incorrect recording of results. Other initial false-positive test results will be likely to be repeated; for example, if there was a biological reason for the initial false-positive test results (such as antibody cross-relativity), the repeat test will probably yield a false-positive result as well. In other words, a person who has had one false-positive test result will have a greater chance of having another false-positive test result.

The population of those who actually do not have the antibody in the unscreened population and the population of those who actually do not have the antibody and are being retested are different: those to be retested all had initial false-positive test results. From this, we can see that on repeat testing a larger percentage of those who actually do not have the antibody will have positive test results because these persons all had one initial false-positive test result. Therefore, the specificity of the repeat EIA will be lower than the specificity of the initial EIA in the unscreened population.

For each confirmatory test, the population to be tested consists of those who were initially EIA-positive from the hypothetical 1 000 000-person blood donor population. From Question 8, the population to have confirmatory testing comprises 20 372 persons, of whom 380 actually have the antibody.

EIA–EIA

Given:

EIA sensitivity 95.0%

EIA specificity 98.0%

See Table 10.21 for numerical results.

Persons are considered to be test-positive only if results of both the initial EIA and the repeat EIA are positive. Because only those with an initial positive EIA were included in Table 10.16, the 761 persons with a repeat positive EIA were positive in

Table 10.21 EIA–EIA testing.

	Disease +	Disease −	Total
Test +	361	400	761
Test −	19	19 592	19,611
	380	19 992	20 372

Table 10.22 EIA–WB testing.

	Disease +	Disease –	Total
Test +	304	2	306
Test –	76	19,990	20 066
	380	19 992	20 372

both the initial and repeat tests. However, of these 761 persons, only 361 actually have the antibody. Therefore, the predictive value positive is 47.4% (361/761).

EIA–WB

Given:

WB sensitivity 80.0%

WB specificity 99.9%

See Table 10.22 for numerical results.

Persons are considered to be test-positive only if results of both the initial EIA and the confirmatory Western blot are positive. Because only those with an initial positive EIA were included in the above table, the 306 persons with a positive Western blot were positive on both tests. Of these 306 persons, 304 actually had the antibody. Therefore, the predictive value positive is 99.3% (304/306).

Note:

Currently, the sequence many blood banks use for notification purposes is EIA–EIA–Western blot (i.e., the original EIA, a repeat EIA, then a Western blot only for those positive on both EIAs). Table 10.23 shows the results of subjecting those blood specimens that are positive on both EIAs to a Western blot test.

EIA–EIA–WB

Given:

WB sensitivity 80.0%

WB specificity 99.9%

See Table 10.23 for numerical results.

Predictive value positive = 289/289 = 100%

Number missed = 400 – 289 = 111

Sensitivity of the entire EIA–EIA–WB sequence = 72%

Specificity of the entire EIA–EIA–WB sequence = 100%, because false positive cell = 0.

Answer 10 From these two examples, we can see that the two most important factors in determining predictive value positive are the prevalence of the disease

Table 10.23 EIA–EIA–WB testing.

	Disease +	Disease –	Total
Test +	289	0	289
Test –	72	400	472
	361	400	761

and the specificity of the test. In the EIA–EIA example, the predictive value positive increased from 1.9% after the initial EIA to 47.4% after the repeat EIA, even though the sensitivity and specificity were the same for both initial and repeat tests. This improvement resulted from the higher prevalence of the antibody in the retested population. For the unscreened population; the prevalence was 0.04%, while for the population being retested, the prevalence was 1.9%. In the EIA–WB example, the predictive value positive after the Western blot test was 99.3%—a marked improvement over repeating the EIA (PVP = 47.4%). This improvement was a result of Western blot's very high specificity (99.9%), even though the sensitivity of the Western blot was much lower than that of the EIA (80 and 98%, respectively).

Answer 11 The criteria to be used in evaluating this screening program include:

1. *Validity*: How well does the test measure what it is supposed to measure? In our premarital program, we are concerned about differentiating those who are infected and those who are not infected. The best scientific evidence to date indicates that those who are exposed to HIV, and hence are antibody-positive, probably remain infected for life. Therefore, an actual antibody-positive person is probably currently infected. It is unknown, however, which infected persons are capable of transmitting the infection.
2. *Reproducibility*: If you repeat the test on the same person, will you get the same result?
3. *Test performance*: What is the yield of the test in terms of sensitivity, specificity, and predictive value?
4. *Cost*: What is the cost of the program?
5. *Follow-up*: Will there be a mechanism to follow up on those with a positive test result?
6. *Acceptance*: Will those who are to be screened accept the program, and will the program be accepted by those performing the follow-up services?
7. *Confidentiality*: Can privacy be maintained?
8. *Public health impact*: Does notification affect behavior?
9. *Prevalence*: Low prevalence inevitably results in many false positives and few true positives.
10. *Feasibility*: What resources and technology are available? What other activities would the screening program displace?
11. *Other benefits*: Source of surveillance data.
12. *Alternatives*: Are there are other programs that would meet the same objectives?
13. *Coverage*: Does the program address those at risk?
14. *Consequences of misclassification*: What are the consequences of being falsely labeled as having the ailment being screened? What are the consequences of being falsely labeled as disease-free?

Answer 12 The costs are:

$3 000 000 Initial screening for all (60 000 × $50 00)

$14 300 Confirmatory testing of those who are initially EIA positive (143 × $100)

Answer 13 This question is intended to provoke discussion. (There is no consensus answer.) However, most would probably not recommend the screening program to the governor. In considering the criteria in Question 11, the screening program

probably meets the criteria of validity, reliability, and yield (high sensitivity, specificity, and predictive value). The program is definitely not cost-effective; the $3 million anticipated cost for this program that would identify 22 antibody-positive persons exceeds the total AIDS budget for most states. The program is likely to be only marginally acceptable to the general population, and there is no proposed mechanism for follow-up of antibody-positive persons. It is also unknown whether notification of antibody-positive persons will cause them to change their sexual practices to reduce the risk of sexual transmission or whether notification will deter them from having children. The program only tests persons at one point in time, shortly before marriage. Therefore, the program would miss persons who have children out of wedlock and those who became antibody-positive after marriage.

CHAPTER 11
The Infectious Disease Process

11.1 The infectious disease process
- Agents
- Reservoirs
  - Symptomatic cases
  - Carriers
  - Animals
  - Inanimate objects
- Portals of entry and exit
- Transmission
  - Mode of transmission
  - Dynamics of transmission
  - Infectious cycles in nature
- Host immunity
  - Types of immunity

11.2 Herd immunity
- What is herd immunity?
- Stemming an outbreak through herd immunity
- Epidemic modeling

Exercises

Review questions

References

> Epidemics never arise from a single cause, but from the interaction of several, at time numerous causes; their strength depending on various influences.
>
> *G. Francke and V. Görttler (1930)*

11.1 The infectious disease process

Studying infectious disease epidemiology is important for two different reasons. First, infectious disease epidemiology provided the original model for the study of disease on a population basis. Many general epidemiologic principles emerged when studying infectious agents and have since been adopted by other fields of epidemiology. For example, the interaction of agent, host, and environmental factors in determining

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

levels of disease in the population was first recognized in an infectious disease context, as was the importance of ordering multiple causal components into causal pathways. The insufficiency of agent-only theories of disease was initially an infectious disease concept.

There are many conceptual similarities between infectious disease epidemiology and chronic disease epidemiology. In fact, many prominent epidemiologists believe division of epidemiology into subspecialites of infectious disease epidemiology and chronic disease epidemiology is arbitrary and detrimental to the discipline (Barrett-Connor, 1979; Stallones, 1980; Susser, 1985).

Second, infectious and parasitic diseases remain a leading cause of morbidity and mortality worldwide (World Health Organization, 1992; National Institute of Allergy and Infectious Disease, 1992). Many infectious diseases have recently emerged (e.g., HIV, hantavirus) or reemerged in virulent forms (e.g., tuberculosis, yellow fever) to imperil the public's health, and the risk of bioterrorism has increased. Therefore, studying infectious disease epidemiology in its own right has taken on added relevance.

As an introduction to infectious disease epidemiology, we consider the following components of the infectious disease process:

- agents
- reservoirs
- portals of entry and exit
- transmission
- host immunity.

Agents

Infections are caused by entry and multiplication of microorganisms and parasites in the body of humans and animals. However, ''infection'' is not synonymous with **infectious disease**, since many infections remain inapparent throughout their course. In addition, the presence of living infectious agents on the exterior of the body or on an article of clothing is not infection, but contamination of a surface.

Infectious agents may be classified according to their size, structure, and physiology. The major categories of infectious disease agents (from structurally largest to smallest) are:

- helminths (parasitic worms);
- fungi and yeast (parasitic lower plants that lack chlorophyll);
- protozoans (minute unicellular organisms often having complex life cycles);
- bacteria (microscopic unicellular organisms capable of independent reproduction);
- rickettsia (microscopic intracellular organisms transmitted by *Ixodes* ticks);
- viruses (submicroscopic infectious agents containing their own genetic material but incapable of multiplication external to a host);
- prions (poorly understood infectious proteins, without discernible nucleic acids, that cause central nervous system infections).

Examples of important infectious diseases from each of these categories are listed in Table 11.1

Table 11.1 Examples of important diseases from each category of infectious agent.

Disease	Agent type	Microbiologic discipline	Comment
Schistosomiasis	Helminth	Parasitology	Causes hundreds of thousands of deaths in circumscribed areas of Asia, Africa, and South America
Cryptococcosis	Fungus	Mycology	Causes problems in immunocompromised patients
Malaria	Protozoan	Protozoology	Continues to affect millions of people in specific areas of Asia, Africa, and South America
Acute bacterial diarrheal diseases	Bacteria	Bacteriology	A leading cause of mortality worldwide, especially troublesome among infants and children in countries with undeveloped economies; various bacterial agents (e.g., *Escherichia coli*, *Shigella sp.*, *Salmonella*)
Typhus fever	Rickettsia	Bacteriology or virology	Louse-borne disease, historically a concomitant of war and famine. Endemic in mountainous regions of Mexico, Central and South America, Central Africa, and numerous countries in Asia
Acquired immunodeficiency syndrome (AIDS)	Virus	Virology	AIDS, caused by the human immunodeficiency virus, is becoming an increasingly important cause of mortality in developing nations
Creutzfeldt–Jakob disease	Prion	Virology	Progressive subacute spongiform infection of the central nervous system; related to ''mad cow disease''

Reservoirs

The **reservoir** of an agent is the normal habitate in which it lives, multiplies, and grows. Without a reservoir, the agent cannot perpetuate itself in nature.

There are many types of reservoirs. These are:

- symptomatic cases
- carriers
  inapparent carriers
  incubatory carriers
  convalescent carriers
- animals (zoonoses)
  direct zoonoses
  cyclozoonoses
  metazoonoses
  saprozoonoses
- inanimate objects
  water
  food
  soil
  air
  fomites.

Symptomatic cases

Symptomatic cases are people with apparent signs of infection. Examples of human diseases in which the primary reservoirs are acute cases include influenza, measles, and smallpox. However, we should not automatically assume that the acutely ill individual always fulfills the role of a reservoir in nature; in many diseases, acute cases might represent biological dead ends in which an essential phase of the agent does not develop or, for some other reason, transmission is disabled. Even when acutely ill individuals are capable of transmitting the agent, they are not necessarily efficient in doing so; acutely ill individuals might be less likely to circulate in the general population of susceptibles and engage in activities necessary for transmission. In fact, it is often the silent carrier that provides the most efficient means of transmission.

Carriers

Carriers are people who harbor the infectious agent, manifest no discernible signs of infection, yet are potential sources of infection. There are three types of carriers:

- inapparent carriers
- incubatory carriers
- convalescent carriers.

Inapparent carriers remain free of the disease throughout the course of infection, yet are still capable of shedding the agent. An example of a disease in which transmission occurs primarily through inapparent carriers is poliomyelitis. For every 100 poliomyelitis infections, only 1 becomes paralyzed, 4 develop nonparalytic disease, and 95 remain disease-free. Nevertheless, all infected individuals may transmit the agent. An additional example of a disease in which inapparent carriers play a crucial role in perpetuating infection is hepatitis A; only 10% of hepatitis-A-infected children demonstrate jaundice, yet fully half are contagious.

Incubatory carriers transmit the agent prior to the onset of disease. Examples of infectious diseases with large incubatory carrier pools include AIDS and hepatitis B. During the long incubatory phase of AIDS, HIV carriers are contagious. Hepatitis B carriers are infectious for an average of 3 months before signs appear.

In the case of **convalescent carriers**, infected persons have recovered from the disease in question but still harbor the agent. A well known case of a convalescent carrier was Mary Mallon—the infamous ''Typhoid Mary.'' Typhoid Mary was free of typhoid symptoms yet continued to harbor and shed the typhoid bacilli throughout her life. She is, perhaps, the world's best known chronic convalescent carrier, having infected at least 53 persons, resulting in three known deaths (Gordon, 1986). However, Typhoid Mary was hardly unique. In general, 1 typhoid patient in 20 continues to shed the infectious agent for at least a year after recovery. Some, like Mary, excrete the typhoid bacilli for life. Convalescent carriers who continue to harbor infection for more than a year are called **chronic carriers**. For some bacterial diseases, incomplete treatment with antibiotics increases the likelihood of the convalescent carrier state. This is why it is important to complete the full course of antibiotic therapy, even after symptoms have abated.

Animals

Zoonoses are infections naturally transmitted between lower vertebrate animals and humans. A less anthropocentric view of zoonoses suggests that they are infections in

which the agent is *shared* between species. Zoonoses constitute a large and diverse group of diseases (over 150 such diseases exist under natural conditions), many of which still cause substantial morbidity and mortality worldwide.

Many zoonotic disease agents have complex life cycles with mandatory intermediate hosts and insect vectors. Accordingly, the following classification scheme for zoonoses has been established:

Direct zoonoses require only a single animal reservoir species to maintain the agent's infectious life cycle. Examples of direct zoonoses are rabies, brucellosis, and trichinosis.

Cyclozoonoses require at least two vertebrate species to complete their infection cycle. Examples of cyclozoonoses are infections with tapeworms and hydatid cysts (larval stages of the *Echinococcus* tapeworm).

Metazoonoses are transmitted to vertebrate hosts by invertebrates, requiring invertebrate vectors or intermediate hosts. Examples of important metazoonoses are schistosomiasis, Lyme disease, arthropod-borne viral diseases (e.g., yellow fever), and plague. Schistosomiasis has a snail intermediate host, Lyme disease is transmitted by an *Ixodes* tick, arthropod-borne viral diseases are transmitted by mosquitoes, and plague is transmitted by a flea.

Saprozoonoses are zoonoses that require inanimate reservoirs in addition to their animal reservoirs. An example of a saprozoonosis is coccidioidomycosis (valley fever), which is caused by a fungus that grows in the soil as a saprophytic mold. It infects humans, cattle, cats, dogs, horses, sheep, wild desert rodents, and other animal species.

Inanimate objects

Some infectious agents are free-living in the environment, growing in inanimate objects such as water, food, soil, air, and other inert substances. Examples of infectious agents with inanimate reservoirs are legionellosis (in which the gram-negative bacillus grows and multiplies in pools of water such as those produced by cooling towers and evaporative condensers), histoplasmosis (a fungal disease with a soil reservoir), and staphylococcal food poisoning (in which the agent multiplies in food, producing toxins capable of causing gastroenteritis).

Portals of entry and exit

For an infectious agent to propagate itself in nature, it must leave one host and enter another. Exit and entry sites for pathogens are called portals. There are six portals in the body:

- respiratory tract (upper and lower);
- conjunctiva (mucous membranes surrounding the eye);
- urogenital tract (urinary tract, sexual genitalia; and organs);
- gastrointestinal tract (upper and lower);
- skin (both intact and broken skin);
- placenta (vertical transmission to offspring).

Blocking an agent's portal can effectively prevent its transmission. Thus, condoms are recommended for the prevention of sexually transmitted diseases and rubber gloves are recommended for the prevention of **nosocomial** (hospital-borne) infections.

Inadvertent transdermal transmission of HIV and hepatitis B can be prevented by exercising sufficient care in the disposal of needles and other sharp clinical and surgical instruments.

In general, agents exhibit preferred portals of entry and exit. For example, tuberculosis bacilli and influenza viruses enter and exit through the respiratory tract, schistosomiasis enters through the skin of humans and exits through the urine or feces (depending on species), and gonorrhea is generally transmitted though the urogenital tract. However, there are some agents that use multiple routes of entry and exit. For example, HIV may enter and exit through the urogenital tract (vagina or penis), gastrointestinal tract (rectal mucosa), skin, and placenta (mother to child).

Transmission

Mode of transmission

Transmission refers to any mechanism by which an infectious agent is spread to another host. In order for an agent to pass from one host to another, the gap between portals must be bridged. Transmission of infection can be accomplished by means of direct and indirect **contact**, by **vectors** (animate objects), and by **vehicles** (inanimate objects). A classification scheme for the modes of transmission is as follows:

- Contact
  direct (requiring physical contact between hosts);
  indirect (contact with relatively fresh bodily fluids or tissue);
  droplets (large infectious particles sprayed from a respiratory portal of an infected host to a susceptible host propelled over a short distance by sneezing or coughing);
  droplet nuclei (small aerosolized particles suspended in air and capable of traveling considerable distances).
- Vectors (animate intermediaries)
  mechanical transmission (no multiplication of the agent in the vector);
  developmental transmission (the infectious organism undergoes a necessary period of development or maturation in the vector);
  propagative transmission (the organism undergoes multiplication in the vector);
  cyclopropagative transmission (the organism multiplies and undergoes development in the vector).
- Vehicles (inanimate intermediates)
  mechanical transmission
  developmental transmission
  propagative transmission
  cyclopropagative transmission.

Examples of diseases transmitted by contact are sexually transmitted diseases, mononucleosis, surgical wound infections, and most respiratory diseases. Examples of vector-borne transmission are malaria (mosquito-borne), Lyme disease (tick-borne), and plague (flea-borne). Examples of diseases transmitted by vehicles are foodborne diseases (e.g., salmonellosis) and waterborne diseases (e.g., cryptosporidiosis).

Dynamics of transmission

Diseases can be transmitted by means of a common vehicle or by serial transfer. **Common vehicle spread** refers to transmission of an agent through a common

source. Examples of common vehicles that may serve this purpose are air, water, food, and drugs. Examples of common vehicle transmission are foodborne disease outbreaks spread by the ingestion of a single contaminated food source or beverage, respiratory disease outbreaks spread by common vehicle transmission through inhalation of air from a contaminated environment (e.g., legionellosis disease), and needle sharing serving as a common vehicle for bloodborne pathogens.

Serial transfer refers to transmission from human to human, human to animal to human, and human to environment to human in sequence. Examples of serially transmitted diseases are measles (spread by the respiratory route from infected to susceptible individuals), sexually transmitted diseases, and any of the diseases requiring person-to-person contact (e.g., AIDS).

Infectious cycles in nature

Many infectious agents have complex **biological cycles**, requiring specific transfers between hosts of different species and within the body of a given host. For example, schistosomiasis (the human blood worm) is acquired from water contaminated with larval forms. The eggs of the worm leave the mammalian host either with the urine or feces (depending on species). Eggs hatch in water, liberating a larval form (miracidium) that enters a suitable freshwater snail host. A different larval phase (cercariae) emerges from the snail and penetrates the human skin while the human host is immersed in a contaminated water source. The cercariae enter the blood stream, are carried to the lungs, and migrate to the liver, where they develop to maturity. The mature worm migrates to the mesenteric and pelvic veins where eggs are deposited and eventually escape to the lumen of the bladder (*Schistosoma haematobium*) or bowel (other *Schistosoma* species). The complex life cycle of *Schistosoma* species is illustrated in Figure 11.1.

Understanding the **natural history of the infectious agent within the host** may also be useful in minimizing the risk of transmission. For example, the infectivity of HIV is determined by its stage of development within the infected host. The natural history of HIV infection is summarized in Figure 11.2. During the acute phase of infection, hosts demonstrate high virus titers and associated high levels of infectiousness, even though an antibody response is lacking (Jacquez *et al*., 1994; Koopman, 1996; Piatak *et al*., 1993; Royce *et al*., 1997). This is followed by a period of low viral titers and, hence, low infectivity. During the latter stages of HIV infection—indicated by symptoms of disease—the host is highly infectious as a result of high viral titers (de Vincenzi, 1994; Lazzarin *et al*., 1991; Lee *et al*., 1996; Royce *et al*., 1997).

Host immunity

Types of immunity

Immunity refers to all factors that alter the likelihood of infection and disease in the host once the agent is encountered. There are two categories of immunity: innate immunity and acquired immunity. **Innate immunity** refers to inborn physical, chemical, cellular, and other physiologic barriers to disease and infection. **Acquired immunity** refers to resistance developed by a host as a result of previous exposure to a natural or artificial pathogen or foreign substance.

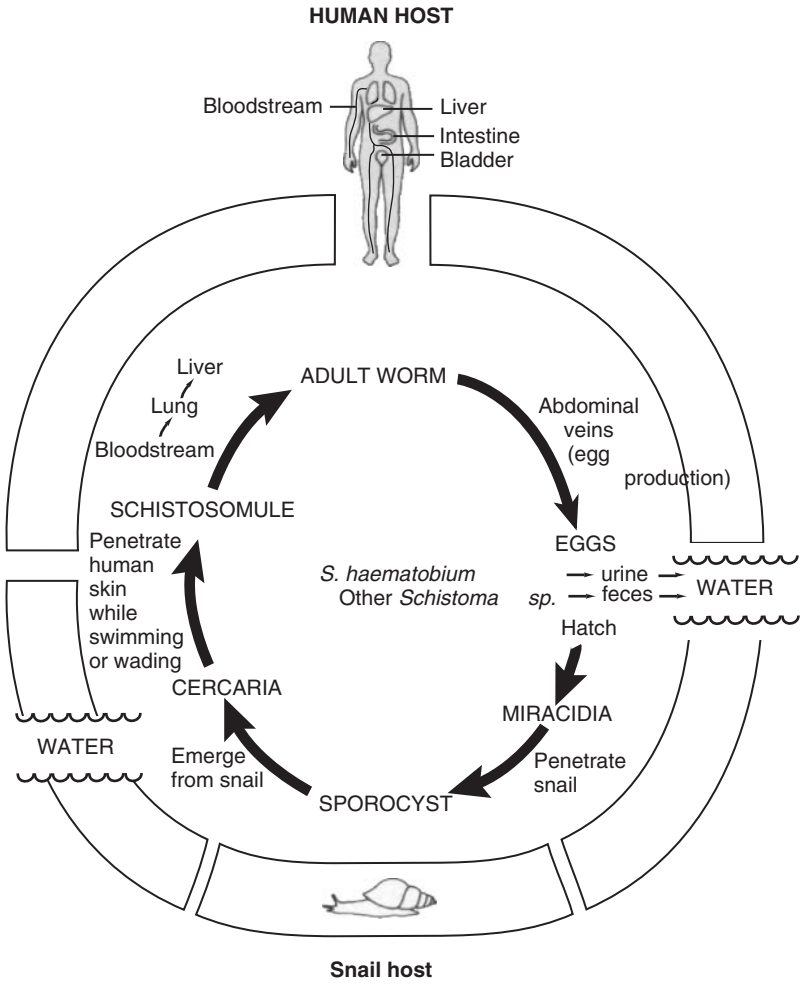


Figure 11.1 Life cycle of *Schistosoma* species (inner circle) and location of the parasite in the host and environment (outer circle).

Innate immunity

Innate immunity includes those barriers that prevent invading pathogens from entering the body and thus establishing themselves within the host. It also includes nonspecific chemical, cellular, and inflammatory bodily reactions that are present from birth.

Examples of innate **physical barriers** to infection are:

- intact skin
- mucosa linings of organs and body cavities
- mucus sheaths on mucosal surfaces
- cilia in the respiratory tract
- cough and gag reflex.

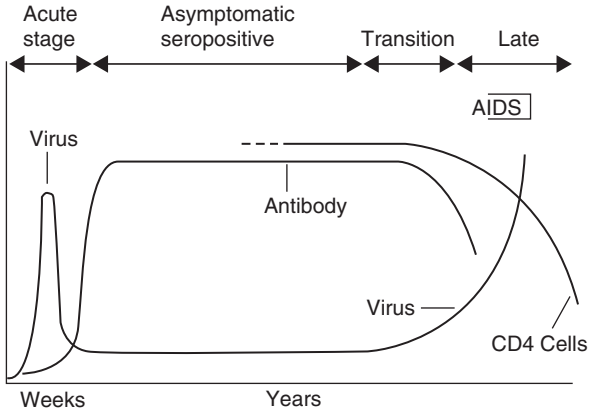


Figure 11.2 Course of HIV infection. (Source: Institute of Medicine, 1988, reprinted with permission from Confronting AIDS: Update 1988. Copyright 1988 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, D.C.)

Examples of innate **chemical barriers** to infection are:

- acidity of the stomach and vagina;
- hydrolytic and proteolytic enzymes in saliva and of the intestines;
- miscellaneous biologically active substances, such as enzymes, lipids, and other molecules (e.g., interferons) that create a hostile environment for invading pathogens.

Examples of innate **cellular and physiologic barriers** to infection are:

- macrophages (large, motile, phagocytic cells found in tissue);
- polymorphonuclear cells (neutrophils and other white blood cells with deeply lobed nuclei capable of chemotaxis, adherence to immune complexes, and phagocytosis);
- reticular endothelial cells (circulating monocytes and stationary phagocytic cells distributed widely throughout the body);
- natural killer cells (cells that release extracellular lytic enzymes);
- inflammation (the body's nonspecific response to injury, including the biologic injury due to infecting pathogens);
- fever.

Innate immunity forms the first line of defense before acquired immunity gets a chance to respond to specific pathogens. These factors are genetically determined but can be modified by host attributes such as age, hormonal status, nutritional status, and physiologic states such as pregnancy and emotional distress.

Acquired immunity

Acquired immunity is the result of a highly specific and evolved response on the part of the host that begins when the host is exposed to a foreign pathogen or substance. Acquired immunity comprises cellular (**immunocytes**) and noncellular (**humoral**) components. There are two types of immunocytes:

- lymphocytes (mononuclear white blood cells found in the lymph, blood, and lymphoid tissue);
- none marrow stem cells (progenitors of other immunocytes).

There are two types of lymphocytes: B lymphocytes and T lymphocytes. **B lymphocytes** (so named because they originate in the bone marrow) synthesize **antibodies** (biochemical proteins that attach themselves to the surface of invading agents), which are secreted into the bloodstream. **T lymphocytes** (so named because they mature in the thymus) help B cells in producing antibodies, neutralize invaders, and regulate other aspects of the immune response through substances called lymphokines. The two forms of immunity—innate and acquired—work closely together to mount the total immunologic response of the host (Figure 11.3).

Immunization

Immunization is the act of acquiring immunity through contact with a foreign substance or agent. There are three types of immunization: active immunization, passive immunization, and adoptive immunization.

Active immunization is a product of the host derived as a result of natural or artificial exposure to **antigens** (foreign proteins associated with the agent or by-products of the agent). **Vaccines** represent artificial exposures that elicit an immune response. Vaccines come in several general forms: **killed vaccines** represent agent proteins incapable of replicating themselves; **modified live vaccines** comprise nonvirulent strains of the agent modified to be nonpathogenic but still capable of stimulating the immune response; **toxoids** are harmless derivatives of microbiologic toxins that stimulate production of antibodies and immunocytes to counter the negative effects of a toxin released by a pathogen or other poisonous source.

Passive immunization is derived from maternal and therapeutic sources. **Maternally derived passive immunity** is acquired through the placenta and colostrum (first milk). **Therapeutic derived immunity** is acquired through the use of immune serums, cytokines, and antitoxins.

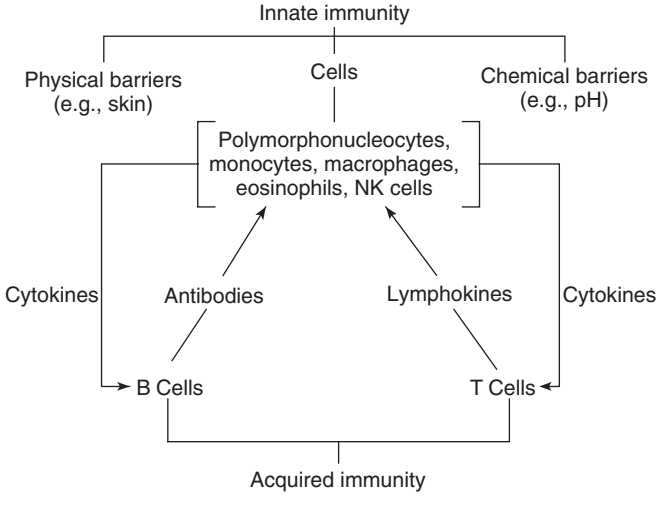


Figure 11.3 Relationships between innate and acquired immunity (Source: Benjamin *et al.*, 1996, p. 38; reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc).

Adoptive immunization involves the transfer of immunocytes from one individual to another. It represents a new technology that holds promise for the future but has few current applications.

11.2 Herd immunity

> The small-pox would be ... sometimes arrested, by vaccination which protected a part of the population.
>
> William Farr (1885, p. 320)

What is herd immunity?

An individual has an immune status that governs his or her susceptibility and, hence, the likelihood of infection once exposed. Similarly, a group has an immune status that governs the susceptibility of the group, and hence the incidence of disease in the ''herd.'' This property of **herd immunity** is defined as the proportion of resistant individuals in the population. As with individual immunity, we speak of innate herd immunity and acquired herd immunity.

Innate herd immunity is the proportion of individuals in the population that are resistant to infection for reason other than prior exposure or immunization. Several examples of innate herd immunity are known. For instance, people with the sickle-cell trait have relatively low parasite levels in the blood when infected with *Plasmodium falciparum* and are thus relatively protected from severe disease (Benenson, 1995). The relative resistance of falciparum malaria in populations with a high prevalence of sickle-cell trait is due to a genetically determined metabolic polymorphism of red blood cells. While this polymorphism is fatal when seen in the homozygous form, and while heterozygosity renders some physiological disadvantages, its overall benefit to survival in endemic *P. falciparum* areas outweighs any such disadvantage. Thus, the presence of the agent in the environment selects for the sickle-cell trait over successive generations. When the selection pressures of falciparum malaria are removed from the population, the frequency of this otherwise disadvantageous trait begins to decline.

Acquired herd immunity is the proportion of individuals in the population resistant to infection as the result of earlier exposure or immunization. The ultimate goal of a vaccination program is to reach an effective level of coverage so that the disease is stopped in its tracks through a process of acquired herd immunity.

Stemming an outbreak through herd immunity

Herd immunity need not be absolute in order to halt the spread of infection through the ''herd.'' When a high percentage of individuals in a population are resistant, transmission may dead-end before reaching remaining susceptible individuals. Figure 11.4 illustrates how this might work. This figure assumes we are dealing with an infection that is transmitted by direct contact. A single case is introduced into the population (cross-hatched). In scenario 1 there is no herd immunity and the agent spreads to all susceptible individuals (infection risk = 100%). In scenario 2 the herd immunity level is 65% (13 of 20) and the agent is blocked after infecting only 2 susceptible

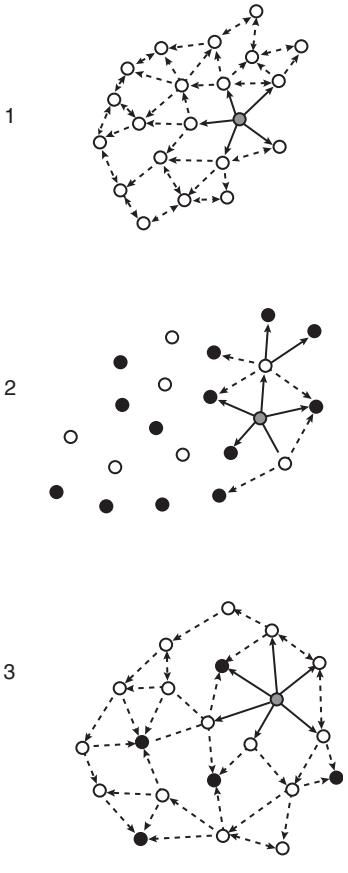


Figure 11.4 Hypothetical spread of contact-transmitted infection upon introduction of a primary case into three different populations: (1) a population with 0% herd immunity, (2) a population with 65% herd immunity, and (3) a population with 25% herd immunity. Key: ◉ primary case, ○ susceptible, • immune, and arrows indicate transmission.

individuals (infection risk = 2/7 = 29%). In scenario 3 the herd immunity level is 25% (5 of 20) and the agent spreads to all remaining susceptible people (infection risk = 100%). Thus, in scenario 2, a herd immunity threshold protects some of those who might otherwise be susceptible. This type of herd immunity stemming threshold depends on the infectivity of the agent and the rate of effective contact. To confer protection to susceptible individuals for a disease that is highly contagious, such as measles, herd immunity must be high (e.g., 95%). In populations where the ''social distance'' between individuals is small, herd immunity rates must be higher still (e.g., 99%) to stem infection (Berger, 1999). With a less contagious agent, such as mumps, a lower frequency of immunity is necessary to prevent outbreaks.

Moderate levels of herd immunity can slow the spread of infection without completely halting its spread. This can have the negative consequence of delaying infection to older ages in those who eventually contract the disease. Since some infectious diseases are more severe or cause more serious consequences when contracted in adulthood (e.g., mumps, chickenpox, hepatitis A, and rubella), a semieffective level of vaccination that affords incomplete herd immunity may decrease the number of infections, but the disease in those who ultimately become ill may have more disastrous effects (Panagiotopoulos *et al*., 1999; Edmunds and Gay, 2000).

Epidemic modeling

Epidemic modeling uses mathematical systems to predict the dynamics of infection, estimate parameters concerning incubation and infection rates, and educate decision makers about vaccination programs. The **Reed–Frost model** is a simple epidemic models in which

$$C_{t+1} = S_t\left(1-q^{Ct}\right)$$

where C_{t+1} is the number of incident cases in time period $t + 1$, S_t is the number of susceptibles in time period t, and $1 - q^{Ct}$ is the probability of having *at least* one effective contact during the interval. An ''effective contact'' is defined as an exposure that would result in infection if one of the individuals was infectious and the other was susceptible; when a susceptible individual has an effective contact with a transmitter, the susceptible person will develop into a case. Thus, the Reed–Frost equation models the course of an epidemic based on the number of susceptible individuals in the population and probability of having an effective contact (''mass action principle'').

Many important assumptions are made in applying the Reed–Frost model (e.g., random mixing of the population, the population is closed to outside contact, conditions of transmission remain constant over time, the infection is transmitted by direct contact only, infectious cases are immune in subsequent time periods). Although the Reed–Frost model can be modified to simulate more realistic assumptions (e.g., two or more open populations with different within- and between-population contacts and random mixing of contagious, immune, and susceptible individuals), the main value of the model today is as a teaching tool, where it is used to demonstrate mass action principles of contagion. More sophisticated mathematical models of transmission have since been developed that incorporate realistic contact structures and social patterns of behaviors (e.g., Longini *et al*., 1988).

Exercises

11.1 Which of the following items *cannot* serve as a portal of entry or exit?
- **(A)** cardiovascular system
- **(B)** skin
- **(C)** respiratory tract
- **(D)** urogenital tract.

11.2 Which of the following is *not* an innate factor of immunity?
- **(A)** gastric acid
- **(B)** cilia in the respiratory tract
- **(C)** antibodies
- **(D)** mucous membranes.

11.3 Toxoids confer:
- **(A)** innate immunity
- **(B)** natural immunity
- **(C)** active immunity
- **(D)** passive immunity.

11.4 Passage of maternal antibodies from mother to child through the placenta confers which type of immunity?
(A) innate immunity
(B) natural immunity
(C) active immunization
(D) passive immunization.

11.5 Which of the following can act as a reservoir?
(A) inanimate objects
(B) animals
(C) carriers
(D) acute cases
(E) only A and C
(F) only A and B
(G) A, B, and C
(H) A, B, C, and D.

11.6 Which type of carrier may result from incomplete treatment with antibiotics?
(A) incubatory carrier
(B) inapparent carrier
(C) convalescent carrier.

11.7 Name the type of carrier that remains free of disease throughout the course of infection.
(A) chronic carrier
(B) inapparent carrier
(C) incubatory carrier
(D) convalescent carrier.

11.8 The transmission of the malaria protozoan through the bite of a mosquito is an example of which mode of transmission?
(A) vehicle borne
(B) direct contact
(C) airborne
(D) vector borne.

11.9 A zoonotic disease requiring only a single animal reservoir and no inanimate reservoir to maintain its life cycle is best classified as a:
(A) direct zoonosis
(B) cyclozoonosis
(C) metazoonosis
(D) saprozoonosis.

Review questions

R.11.1 How does infection differ from contamination?

R.11.2 How does infection differ from infectious disease?

R.11.3 List the types of infectious disease agents, from largest to smallest.

R.11.4 Name the four types of reservoirs.

R.11.5 How does a case differ from a carrier?

R.11.6 What do you call an infectious disease in which the agent is shared by humans and other species?

R.11.7 List types of objects that can serve as reservoirs.

R.11.8 How does a cyclozoonosis differ from a direct zoonosis?

R.11.9 How does a metazoonosis differ from a direct zoonosis?

R.11.10 What does the prefix sapro- mean?

R.11.11 List the six portals for infectious agents.

R.11.12 How does direct contact transmission differ from indirect contact transmission?

R.11.13 How does droplet transmission differ from droplet nuclei transmission?

R.11.14 How does a vector differ from a vehicle?

R.11.15 How does mechanical transmission differ from developmental transmission?

R.11.16 How does developmental transmission differ from propagative transmission?

R.11.17 Was the Golden Square ''Broad Street pump'' cholera outbreak of 1854 (Chapter 1) a common-vehicle spread or serial spread outbreak?

R.11.18 Does the common cold spread by a common-vehicle or serially?

R.11.19 How does knowledge of an agent's biological cycle help with its control?

R.11.20 How does understanding the natural history of a disease help with its control?

R.11.21 True of false? Nonspecific inflammatory responses in the host (e.g., fever) are innate forms of immunity.

R.11.22 How does passive immunization differ active immunization?

R.11.23 Identify two forms of passive immunization.

R.11.24 Intentional exposure to an antigen in order to initiate an immune response is called _____.

R.11.25 Generally, do killed or modified live vaccines produce more sustained immune responses? Explain.

R.11.26 Cell mediated immune responses are mediated through biochemical cell proteins called _____.

R.11.27 What type of immunocyte produces antibodies?

R.11.28 HIV selectively attacks what type of immunocyte?

R.11.29 B and T cells are types of _____.

R.11.30 What are two different reasons to study the infectious disease process?

R.11.31 Define herd immunity.

R.11.32 How does innate herd immunity differ from acquired herd immunity?

R.11.33 Describe how herd immunity works.

References

Barrett-Connor, E. (1979) Infectious and chronic disease epidemiology: Separate and unequal? *American Journal of Epidemiology*, **109**, 245–249.

Benenson, A. S. (ed.) (1995) *Control of Communicable Disease Manual*, 16th edn, American Public Health Association, Washington, DC.

Benjamini, E., Sunshine, G., and Leskowitz, S. (1996) *Immunology. A Short Course*, 3rd edn, Wiley-Liss, New York.

Berger, A. (1999) How does herd immunity work? *British Medical Journal*, **319**, 1466–1467.

de Vincenzi, I. (1994) A longitudinal study of human immunodeficiency virus transmission by heterosexual partners. *New England Journal of Medicine*, **331**, 341–346.

Edmunds, W.J., and Gay, N.J. (2000) Health professionals do not understand mathematical models. *British Medical Journal*, **320** (7234), 581a.

Farr, W. (1885) *Vital Statistics*: *A Memorial Volume of Selections from the Reports and Writings of William Farr*, (*1975 Reprint*). Offices of the Sanitary Institute, Metuchen, NJ.

Francke, G., and Görttler, V (1930) *Allgemeine Epidemiologie der Tierseuchen*, Ferdinand Enke, Stuttgart.

Gordon, R. (1986) *Great Medical Disasters*, Dorset Press, New York, chap. **16**.

Jacquez, J.A., Koopman, S.J., Simon, C.P., and Longini, I.M., Jr., (1994) Role of primary infection in epidemics of HIV infection in gay cohorts. *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology*, **7**, 1169–1184.

Koopman, J. (1996) Emerging objectives and methods in epidemiology. *American Journal of Public Health*, **86**, 630–632.

Lazzarin, A., Saracco, A., Musicco, M., and Nicolosi, A. (1991) Man-to-woman sexual transmission of the human immunodeficiency virus: Risk factors related to sexual behavior, man's infectiousness, and woman's susceptibility. *Archives of Internal Medicine*, **151**, 2411–2416.

Lee, T.H., Sakahara, N., Fiebig, E., Busch, M.P., O'Brien, T.R., and Herman, S.E. (1996) Correlation of HIV-1 RNA levels in plasma and heterosexual transmission of HIV-1 from infected transfusion recipients. *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology*, **12**, 427–428.

Longini, I.M., Jr.,, Koopman, J.S., Haber, M., and Cotsonis, G.A. (1988) Statistical inference for infectious diseases. Risk-specific household and community transmission parameters. *American Journal of Epidemiology*, **128**, 845–859.

National Institute of Allergy and Infectious Disease (1992) *Report of the Task Force on Microbiology and Infectious Disease*, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, NIH Publication 92–3320.

Panagiotopoulos, T., Antoniadou, I., Valassi-Adam, E., and Berger, A. (1999) Increase in congenital rubella occurrence after immunisation in Greece: Retrospective survey and systematic review. *British Medical Journal*, **319**, 1462–1467.

Piatak, M., Jr.,, Saag, M.S., Yang, L.C., *et al*. (1993) High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science*, **259**, 1749–1754.

Royce, R.A., Sena, A., Cates, W., and Cohen, M.S. (1997) Sexual transmission of HIV. *New England Journal of Medicine*, **336**, 1072–1078.

Stallones, R.A. (1980) To advance epidemiology. *Annual Review of Public Health*, **1**, 69–82.

Susser, M. (1985) Epidemiology in the United States after World War II: The evolution of technique. *Epidemiologic Reviews*, **7**, 147–177.

World Health Organization (1992) *Global Health Situations and Projections, Estimates 1992*. WHO; Geneva.

CHAPTER 12
Outbreak Investigation

12.1 Background
- Initial detection of outbreaks
- Goals and methods of outbreak investigations

12.2 CDC prescribed investigatory steps
- Step 1: Prepare for field work
- Step 2: Establish the existence of an outbreak
- Steps 3 and 4: Verify diagnoses of cases and search for additional cases
- Step 5: Conduct descriptive epidemiologic studies
  - Time
  - Place
  - Person
- Step 6: Develop hypotheses
- Steps 7 and 8: Evaluate hypotheses; as necessary, reconsider or refine hypotheses and conduct additional studies
- Step 9: Implement control and prevention measures
- Step 10: Communicate findings

Review questions

References

Chapter addendum 1 (case study)
- Drug–disease outbreak
  - Background
  - Preparatory research
  - How hGH is prepared
  - Source at the agent
  - More cases
- Answers to case study: a drug–disease outbreak

References—a drug–disease outbreak

Chapter addendum 2 (case study)
- Food borne outbreal in Rhynedale, California
  - Comment
  - Background
  - The Locale
  - The church supper
  - The involved group
  - Materials and methods
  - Data
- Answers to case study: food-borne disease outbreak

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

> The essentials of the epidemiological approach in determining the specific etiologic agent of a disease are demonstrated in their simplest form by the investigation of the etiology of a food poisoning outbreak.
>
> *A.M. Lilienfeld (1976, p. 15)*

12.1 Background

Initial detection of outbreaks

An outbreak is an epidemic or upsurge of cases in a defined geographic region or easily defined subpopulation. Outbreaks come to the attention of public health agencies in two primary ways: either through surveillance systems or by notification of public health departments by citizens or health care providers.

Epidemiologic surveillance systems are organizations and structures set in place to collect and analyze outcome-specific health data. Surveillance systems are distinguished by their practicability, uniformity, and rapidity at which they receive and process information. As a result, information from surveillance systems is often limited in scope, and often requiring supplementation. (See Section 4.1 for additional background on surveillance systems). The other way to become aware of an outbreak is by **direct notification** by an affected individual, their family, or their care givers. See Illustrative Example 4.1 for a discussion about the initial detection of HIV/AIDS.

The decision whether to mount an investigation of an apparent outbreak is based on the numerous factors, including: (a) the ability to confirm that the observed number of cases is significantly greater than expected, (b) the scale and severity of the outbreak, (c) whether the outbreak disproportionally affects an identifiable subgroup, (d) the potential for spread, (e) political and public relations considerations, and (f) availability of resources. Ultimately, the decision comes down to the authorities at the local health department made in consultation with the Centers for Disease Control and Prevention or other comparable agency outside of the United States.

The responsibility of investigating outbreaks usually falls on the shoulders of the local county or city health department. However, if the investigation requires additional resources, attracts substantial public concern, or is associated with a high attack rate and serious complications (hospitalization or death), state and federal agencies are called in. Outbreaks of national importance are investigated by the CDC. In fulfillment of this responsibility, the CDC has prepared excellent outbreak investigation training materials. Much of the remaining discussion in this chapter is based on these materials.

Goals and methods of outbreak investigations

Outbreak investigations have both diagnostic and directed action components. The objectives of the investigation may include:

- To assess the range and extent of the outbreak.
- To reduce the number of cases associated with the outbreak by identifying and eliminating the source of the problem.
- To identify new disease syndromes.
- To identify new causes of known disease syndromes.

- To assess the efficacy of currently employed prevention strategies.
- To address liability concerns.
- To train epidemiologists.
- To provide for good public relations and educate the public.

To successfully conduct any type of epidemiologic investigation, Evans (1982, p. 7) offers these helpful actions:

1. Define the problem, confirm diagnoses, and show that an outbreak has truly occurred.
2. Describe the outbreak clinically (based on signs, symptoms, and tests) and epidemiologically (based on by person, place, and time variables). Note common features and exceptions. Discuss possible means of transmission. Calculate rates of infection, disease and death.
3. Formulate hypotheses about cause, source of infection, method of contamination and spread, and possible control mechanisms.
4. Test causal hypotheses with epidemiologic, laboratory, and environmental investigations.
5. Draw conclusions and devise practical control solutions.

The CDC (1992a) prescribes these specific steps:

1. Prepare for field work.
2. Establish the existence of an outbreak.
3. Verify diagnoses of cases.
4. Establish a case definition and search for additional cases.
5. Conduct descriptive epidemiologic studies.
6. Develop hypotheses.
7. Evaluate hypotheses.
8. As necessary, reconsider or refine hypotheses and conduct additional studies.
9. Implement control and prevention measures.
10. Communicate findings.

The CDC prescribed steps are discussed in the next section.

12.2 CDC prescribed investigatory steps

Step 1: Prepare for field work

Preparation for an investigation includes completing the administrative and personal measures required to begin the inquiry. Travel preparations must be made, supplies and equipment readied, knowledge updated, and administrative and scientific contacts established. Investigators must have a clear understanding of their role in the field and must know the chain of authority involved in the process.

Step 2: Establish the existence of an outbreak

The task of verifying an outbreak is made simple if a common cause can be identified (as might be expected with food-borne illnesses). When this is the case, mechanisms of transmission and means of control will be known, allowing for routine and rapid completion of the investigation.

However, when a common cause cannot be initially identified, the suspected outbreak must be first confirmed. In so doing, the first step in establishing the existence of an outbreak is to confirm that the reported cases actually have the disease that is being reported. After confirming each case using standard diagnostic criteria, the observed rate of occurrence is compared with that which is normally expected. Rates in the absence of an outbreak can often be gleaned from national surveys, special registries (e.g., cancer or birth defect registries), data from neighboring states, and from the published literature. In addressing this issue, the epidemiologist compares the observed number of cases with that which is expected under normal conditions. The comparison must account for random fluctuations in occurrence (see Sections 19.2 and 19.3). It must also account for seasonal variation (e.g., see Figure 4.8) and other phenomena that could conceivably increase the reported number of cases without a concommitent true increase in occurrence in the population. This includes:

1 **Change in the reporting practices or case definition**. A change in the reporting procedure or case definition may explain the increase in reported cases without a concomitant increase in the rate of disease in the population.
2 **A change in population size**. A sudden increase in population size, such as might occur in a resort area, college town, or farming area with migrant labor, could explain the apparent increase.
3 **Diagnostic suspicion bias**. Increased diagnostic sensitivity such as might occur with improved diagnostic procedures, screening campaigns, or a new physician or infection control nurse in town could explain the increase.
4 **Publicity bias**: Publicity such as might occur when media attention stimulates the reporting of cases that would have previously gone unnoticed, could explain the apparent increase.

Steps 3 and 4: Verify diagnoses of cases and search for additional cases

If the existance of the outbreak is verified, the next task is to review the existing cases and search for additional cases using a standardized case definition. The **case definition** is the set of standardized criteria used to decide whether an individual should be classified as having the disease in question or not.

The investigation team searches for previously unidentified cases in local hospitals and clinics that are likely to treat cases. In addition, they may screen data in clinical laboratories that are likely to diagnose cases. It often proves useful to directly question those individuals who might treat or encounter the disease. For example, in studying a disease of the blood, the investigator might query hematologists and laboratory personnel who treat or study the disease; in studying neoplastic diseases, the investigator questions oncologists, cancer clinics, cancer support groups, and other people likely to encounter prospective cases. Because direct inquiry may require a fair amount of walking about, it has traditionally been called ''**shoe-leather**'' **epidemiology**. Additional cases may be discovered by issuing a plea for reports through a media appeal or by direct requests for information to physicians. Note, however, that blanket requests such as these may elicit duplicate reports, false-positives, reports of old cases irrelevant to the current outbreak, and other dubious information.

Step 5: Conduct descriptive epidemiologic studies

Descriptive epidemiology is used to explore the general pattern of disease in the affected population. Descriptive epidemiology has the following objectives:

- to learn about the range and extent of the outbreak;
- to assess the possible source of exposure, mode of transmission, incubation period, environmental contributors, host risk factors, and agent characteristics;
- to generate hypotheses about the outbreak.

To begin the process of describing the outbreak, the following information is collected:

- case identification information (name, address, telephone number, and other information that will allow investigators to contact the subjects for notification or follow-up purposes);
- demographic information (age, sex, race, occupation, and other ''person'' factors that allow for the description of rates);
- clinical information (time of disease onset, time of exposure to the etiologic agent, signs, symptoms, and test results as are relevant to the case definition);
- risk factor information (relevant exposures and extraneous factors that might influence the occurrence of disease, specific items must be tailored to the disease in question);
- reporter information (to allow for further questioning and follow-up, if needed);
- denominator data (census and *ad hoc* information that might provide reasonable estimates of denominators for prevalence and incidence calculations).

After the data are entered into a database and assessedf for quality and completeness, the investigator describes the outbreak according to **epidemiologic variables** of time, place, and person. Although principles of descriptive epidemiology have already been considered in Chapter 4, a brief review relevant to outbreak investigation is in order.

Time

An important component of the investigation is the **epidemic curve**. The *y* axis of an epidemic curve represents the number (or percentage) of cases. The *x* axis represents a time line. When drawing the curve, the *x* axis should begin before the epidemic period and extend to the period after the epidemic is over.

Epidemic curves provide pictorial insights into the source of exposure, the nature of the epidemic (e.g., whether it is a point source epidemic or propagating epidemic), and the future course of the epidemic.

The incubation period of an agent is the time between exposure to the agent and appearance of first signs or symptoms of disease. This period varies considerably by the type of agent, level of exposure, and susceptibility of the host. If the probable time of exposure to the agent is known, the incubation period can be summarized with descriptive statistics, such as the minimum, maximum, and average. The average incubation can be expressed as an arithmetic mean, geometric mean, median, and/or mode. Knowledge of the range and average incubation period is often helpful in identifying the type of agent and its source.

Figure 12.1 exhibits an epidemic curve for an outbreak of hepatitis B in which exposure was from drinking uncholorinated water at a school. Because the exact date

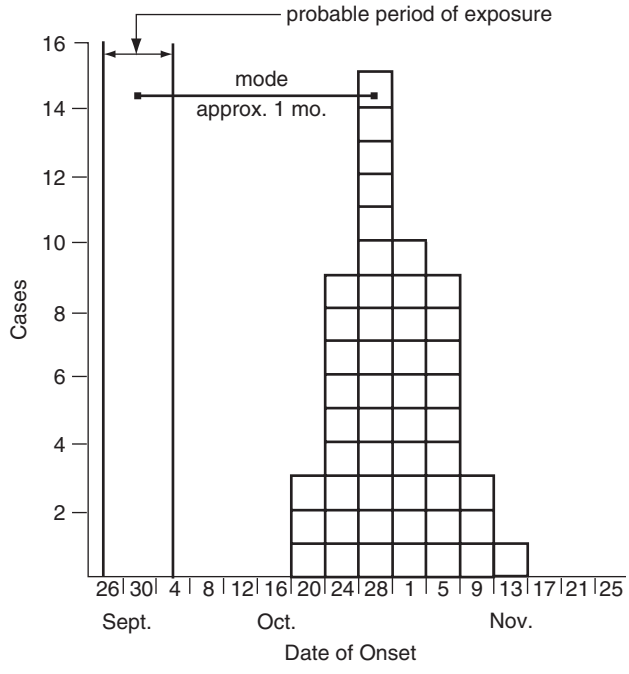


Figure 12.1 Epidemic curve, hepatitis A outbreak, Colbert County, Alabama, 1972. (Source: CDC, 1992a, p. 367.)

of exposure could not be determined, the investigators were able to back-calculate the most likely period of exposure based on the typical one month incubation period of the disease. This yielded a date that was consistent with the period during which water supply at the school was not cholorinated.

Epidemic curves are also useful for predicting the future temporal course of an epidemic. For example, Figure 12.2 shows an epidemic curve for famous 1854 Golden Square (Broad Street pump) cholera epidemic investigated by John Snow (see Section 1.4). Ironically, this graph suggests that removal of the handle from the Broad Street pump had little to do with ending this notorious outbreak, thus contradicting the folklore surrounding this decisive (and largely symbolic) act. Sir A. Bradford Hill (1955) eloquently describes the course of events:

> Though conceivably there might have been a second peak in the curve, and though almost certainly some more deaths would have occurred if the pump handle had remained *in situ*, it is clear that the end of the epidemic was not dramatically determined by its removal. (p. 1010)

John Snow (1855), himself, recognized that the epidemic might have burned itself out before the pump handle was removed.

> . . . but the attacks had so far diminished before the use of the water was stopped, that it is impossible to decide whether the well still contained the cholera poison in an active state, or whether, from some cause, the water had become free from it. (pp. 51–52)

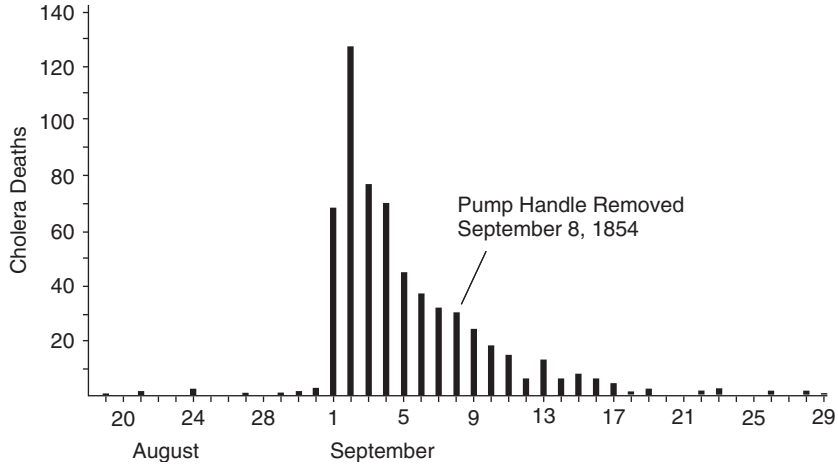


Figure 12.2 Epidemic curve. London cholera epidemic of 1854 (data from Snow, 1855, p. 49).

The shape of the epidemic curve is useful in determining the type of epidemic in question. **Point-source epidemics** are caused by exposure to the agent from a single source over a brief period of time. When this is the case, the epidemic exhibits a sudden rise followed by a rapid falloff (Figure 4.6c). **Propagating epidemics** depend on serial propagation from host to host, or possibly continuous exposure from a single source, and thus exhibit a plateau or continual rise in the number of cases (Figure 4.6d).

Place

Mapping cases according to their place of origin may provide supporting evidence about transmission of the agent responsible for the outbreak. Epidemic maps may take the form of simple dot maps indicating location of origin or maps or area-specific rates.

Dot maps serve to document the geographic extent of the problem and can provide evidence of clustering. Snow's celebrated map of clustering of cholera deaths around the Broad Street pump (Figure 1.13) provides a well-known historical example. By combining the mapping of cases with other sources of information, John Snow was able to support his theory of the waterborne transmission of cholera.

When the populations in the areas being compared are of unequal size, dot maps can be misleading. To compensate for this inherent weakness of dot maps, the epidemiologist will map **rates by region**. One such map of an Ebola virus outbreak is shown as Figure 12.3. This figure displays Ebola attack rates per 100 inhabitants in the epidemic zone. The highest attack rates are centered around Yambuku, Zaire, the town where the mission hospital was located. Decreasing attack rates with increasing distance from the hospital supported a hypothesis of iatraogenic spread of the agent.

Person

Description of disease rates by **person variables** is useful in identifying high-risk groups. Examples of person factors relevant to outbreak investigation include

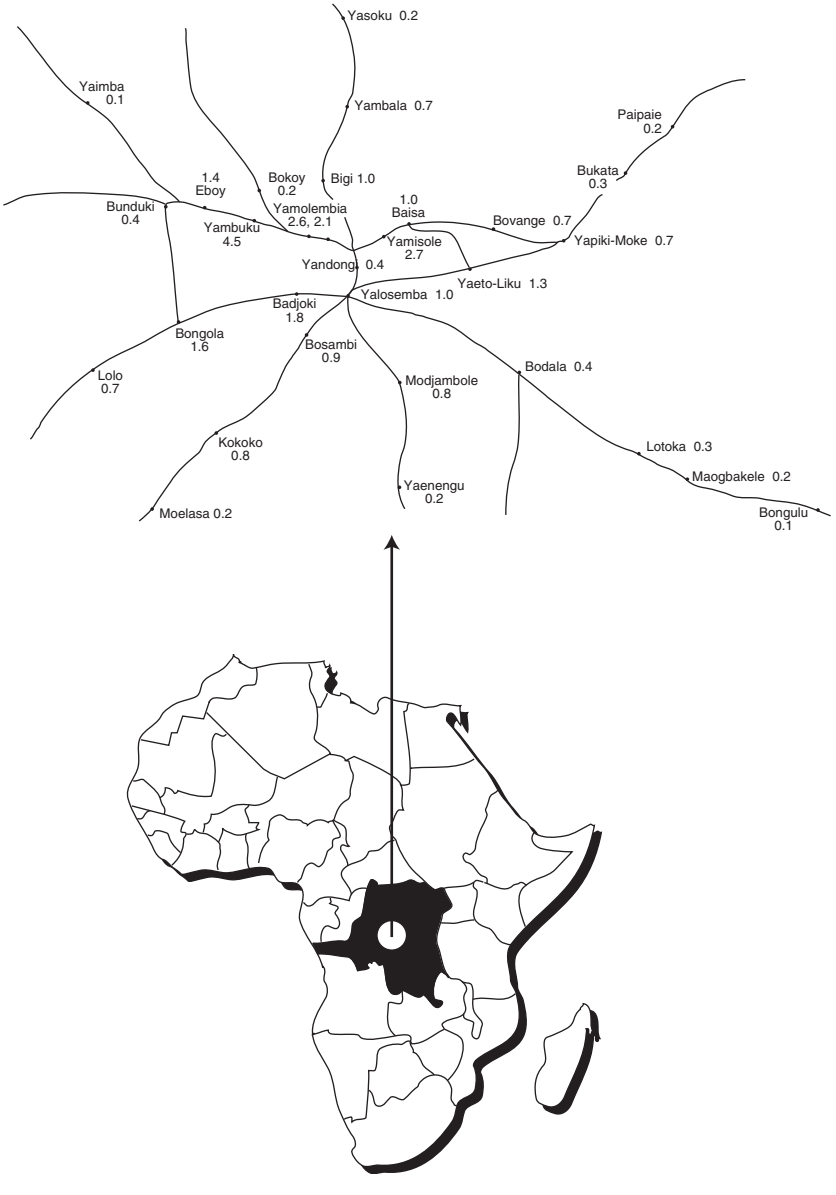


Figure 12.3 Ebola virus epidemic zone, Zaire, 1976. (Source: Adapted from CDC, 1992b, p. 10.)

demographic characteristics (age, sex, ethnicity), personal activities and practices (occupation, customs, leisure activities, religious activities, knowledge, attitudes, and beliefs), genetic predispositions, physiologic states (pregnancy, parity, distress, nutritional status), concurrent diseases, immune status, and marital status. At minimum, the frequency of disease should be described by age and sex. Other analyses according to person variables cater to the type of disease being investigated. For example, when investigating AIDS, the epidemiologist is interested in describing disease rates according

to sexual practices, intravenous drug use, and exposures to blood transfusions and other biological products of human origin.

Step 6: Develop hypotheses

A **hypothesis** is a tentative explanation that accounts for a set of facts and that can be tested by further investigation. In the investigation of outbreaks, hypotheses should address the most likely source of exposure to the etiologic agent, the means of transmission, and the next steps in the investigation and future control measures.

Hypothesis generation involves a scientific knowledge of the facts and a bit of intuition. It begins when the first clues that an epidemic might exist come to light and continues until the investigation is complete. Hypothesis development requires an understanding of the disease process and population at risk. It is supported by discussions with patients, health-care providers, local public health officials, community activists, and other interested parties, and should include the review of all relevant clinical, epidemiologic, and laboratory information. In generating and developing hypotheses one should consider what is generally known about the disease, relevant clinical and laboratory findings, what patients say about the disease, and the the descriptive epidemiologic findings

Table 12.1 is a checklist that may be used when generating and developing hypotheses about outbreaks. When generating hypotheses, we search for common characteristics and notable exceptions.

Table 12.1 Hypothesis-generating checklist.

1.	Review what is known about the disease itself:	
	Agent	Reservoir
	Mechanisms of transmission	Natural history of disease
	Clinical spectrum	Pathogenic mechanisms
	Known risk factors	Ecology of the agent
2.	Study clinical and laboratory findings:	
	Review clinical and laboratory records (check and confirm accuracy)	
	Determine if specialized lab work is necessary (e.g., DNA ''fingerprinting'')	
	Describe frequency of symptoms, signs, and test results among cases	
3.	Consider what patients and caregivers say:	
	Determine potentially relevant exposures	Hear what they think about cause
	Gain additional insights into clinical features	See if they are aware of other cases
	Determine commonalities and differences in cases	
4.	Review descriptive epidemiology:	
	Epidemic curve and pattern	Geographic distribution
	Incubation period statistics	Significant host risk factors
	Events occurring around the most likely period of exposure for each case	
5.	Ruminate facts:	
	Deduction	Intuition
	Analogy	Coherence
	Credibility of sources	Quality of information
	Missing keys and explanations	Exceptions and outliers

When searching for common characteristics, the objective is to search for specific exposures that have the **strongest association** with disease. This is normally done by calculating risk ratios or odds ratios. The exposure associated with the largest risk ratio or odds ratio is a prime suspect as the sources of the agent.

Important clues may also come from investigating **notable exceptions**. For example, we might investigate why certain people who were exposed to the putative agent did not become ill, and why apparently unexposed people did develope the illness. Such ''outliers'' can provide clues about the source of infection and mode of transmission. John Snow, in his classic 19th-century cholera investigations, used this technique repeatedly. For example, he pointed out the relative absence of fatal cholera cases in brewery workers living near the epidemic's center (see Figure 1.13) and attributed this deficiency to avoidance of pump water. (The proprietor of the brewery believed his workers did not drink water at all and most certainly did not obtain water from the pump on the street.) Snow also noted a fatal case in a 59-year-old widow living outside the epidemic area and traced this to water transported from the pump. These notable exceptions provided strong clues in support of the waterborne theory of cholera transmission.

Steps 7 and 8: Evaluate hypotheses; as necessary, reconsider or refine hypotheses and conduct additional studies

Hypotheses developed in step 6 are re-examined, refined, and tested throughout the investigation. The process is iterative, cyclic, and self-correcting, requiring continual hypothesis refinement and testing. The usefulness of the epidemiologic investigation is often dictated by the clarity and quality of its underying hypotheses.

Causal hyoptheses can be tested using qualitative or quantitative methods, depending on underlying circumstances. Here is an example in which the causal hypothesis was solved with **qualitative methods** (CDC, 1992a, p. 375):

> In an outbreak of hypervitaminosis D that occurred in Massachusetts in 1991, it was found that all the case-patients drank milk delivered to their homes by a local dairy. Therefore, investigators hypothesized that the dairy was the source and the milk was the vehicle. When they visited he dairy, they quickly recognized that the dairy was inadvertently adding far more than the recommended dose of vitamin D to the milk. No analytic epidemiology was really necessary to evaluate the basic hypotheses in this setting.

In other instance, **quantitative epidemiologic investigations** will be necessary to draw inferences about the etiology of the outbreak and source of exposure. Analytic studies may take the form cohort or case–control studies. The choice of a study design depends on whether the outbreak is ongoing or has been resolved, the availability of resources, past experience of the investigator, the size of the population at risk, the prevalence of the exposure, and the incidence of the disease. In general, small, well-circumscribed outbreaks in which the incidence of disease is high are well suited for cohort study. In contrast, outbreaks in large, poorly circumscribed populations in which the disease is rare may be better suited for case–control methods.

Laboratory and environmental studies are used to isolate causal agents from cases and from the environment. Environmental and sanitary conditions should be

studied to help explain why the outbreak occurred in the first place and what might prevent it from happening again. Special laboratory tests (e.g., DNA, chemical, or immunologic fingerprinting) can occasionally be used to link the agent isolated from patients to specific environmental sites. When available, laboratory evidence can ''clinch the findings'' established by the epidemiologic investigation. Note, however, that since many outbreaks are investigated after the fact, collection of specimens may be precluded.

Step 9: Implement control and prevention measures

Two of the main objectives of outbreak investigation are to (a) bring the current epidemic to a halt and (b) to prevent future occurrences. Elements of disease control may be directed toward agent, host, or environmental factors. In foodborne outbreaks, for example, it is important to identify the initial point of contaimination, if possible. In some instances, this may require tracking the agent to its initial agricultural source. Remaining contaminated food should be discarded after specimens are collected for laboratory investigation. Food preparers should be educated on proper handling of food in terms of refrigeration, cooking, cooling, storage and serving techniques in order to prevent future occurrences.

Table 12.2 The ''what, why, when, how, where, and who'' of outbreak reporting.

What: oral briefings
Why: to disseminate information and defend conclusions and recommendations, to promote good public relations, and to allow for constructive criticism
When: at the beginning and end of the investigation and whenever information for prevention and control comes to light
How: use scientifically objective language (avoid emotional terms), consider the audience (many people may not be epidemiologists), and explain epidemiologic principles and methods (avoid jargon)
Where: the appropriate venue is dictated by the audience; presentations should be given in the locality affected by the outbreak and at the sponsoring agency; findings can also be presented at regional and national professional conferences
Who: audience may vary but should include local, state, and federal authorities and people responsible for control and prevention measures
What: written reports
Why: to document the investigation, to disseminate information and defend conclusions and recommendations, to promote good professional relations, to increase credibility of the work, to allow for constructive criticism, to prevent future occurrences, and to add to the public health information base
When: at the conclusion of the investigation
How: use standard scientific reporting format with introduction, methods, results, discussion, (± recommendations); sponsoring agency may have additional reporting requirements
Where: internal documents should be filed with the local health department and all supporting agencies; if appropriate, a manuscript should be submitted to a general or discipline-specific peer-reviewed journal for publication
Who: audiences may vary but might include epidemiologists in training, field epidemiologists, and researchers in the discipline

Step 10: Communicate findings

The investigation is not complete until the results are disseminated to the public and the profession. Findings should be reported to initial informants, those involved in the investigation, local, state, and federal public health agencies, and the community of people affected by the outbreak. This is done in the form of oral briefings and written reports. Examples of the ''what, why, when, how, where, and who'' of reporting are summarized in Table 12.2.

Review questions

R.12.1 What federal agency in the United States is primarily responsible for investigating outbreaks of national importance?

R.12.2 Multiple choice (M/C): Select the best response. The *x* axis of an epidemic curve represents: (a) the number of cases; (b) time since exposure or some other type of time line; (c) the percentage of cases; (d) responses a and c; or (e) all of the above.

R.12.3 M/C: The *y* axis of an epidemic curve represents: (a) the number of cases; (b) time since exposure; (c) the percentage of cases; (d) responses a and c; or (e) all of the above.

R.12.4 M/C. A notable exception to the observed pattern of occurrence is a(*n*): (a) outlier; (b) case; (c) control; and (d) exposure.

R.12.5 How do outbreaks come to the attention of public health authorities?

R.12.6 What is an epidemiologic surveillance system?

R.12.7 List four different phenomena that could cause an increase in the number of cases reported to a system without there being a true increase in the rate of occurrence.

R.12.8 The decision whether or not to investigate an outbreak depends on many factors. List several.

R.12.9 Descriptive epidemiology traditionally describes the occurrence of a condition according to three types of ''epidemiologic variables''. Name these.

R.12.10 One of the first things to do when investigating an outbreak is to confirm the _____ of prospective cases.

R.12.11 The observed number of cases is compared with the _____ number to confirm the occurrence of an epidemic.

R.12.12 What does an epidemiologist mean when they refer to a ''case definition''?

R.12.13 True or false? John Snow recognized that the outbreak of cholera in the Golden Square area that was associated with the Broad Street pump might have been burning itself out before the pump handle was removed.

R.12.14 Define ''incubation period.''

R.12.15 What is a hypothesis?

R.12.16 True or false? Hypothesis development requires an understanding of the disease process and population at risk

R.12.17 When communicating the findings of an outbreak investigation, we address the 'who, what, where, when, _____, and _____ of what we discovered.

References

Centers for Disease Control and Prevention (CDC) (1992a) *Principles of Epidemiology: Self-study Course 3030-G*, 2nd edn, U.S. Department of Health and Human Services, Atlanta. GA.

Centers for Disease Control and Prevention (CDC) (1992b) *An Outbreak of Hemorrhagic Fever in Africa, 1992 EIS Course*, Association of Teachers of Preventive Medicine, Washington, DC.

Evans, A.S. (1982) Epidemiological concepts and methods, in *Viral Infections of Humans. Epidemiology and Control* (ed. A.S. Evans) Plenum Medical Book Company; New York, pp. 3–42.

Hill, A.B. (1955) Snow—an appreciation. *Proceedings of the Royal Society of Medicine*, **48**, 1008–1012.

Lilienfeld, A.M. (1976) *Foundation of Epidemiology*, Oxford University Press, New York.

Snow, J. (1936). *Snow on Cholera*, The Commonwealth Fund, New York. (Originally published as *On the Mode of Communication of Cholera* in 1855.)

Chapter addendum 1 (case study)

Drug–disease outbreak

Background

As your first assignment working for the U.S. Food and Drug Administration (FDA), you are to act as an epidemiologic consultant to the division that reviews and approves the use of endocrinologic and metabolic drugs. One day in late February (1985), a medical officer from the reviewing division comes to you with an unusual problem and asks for your help. The medical officer has just received a report of the death of a 20-year old man with Creutzfeldt–Jakob disease (CJD). It is noted that this case received human growth hormone (hGH) for 13 years as a child, between the ages of 3 and 16 (from 1966 to 1980). You note that hGH is used to prevent pituitary dwarfism when given during the growth years.

Question 1 What is your first step in investigating this case? What is your reaction to this single case of CJD?

Question 2 What additional information do you need to evaluate the relevance of this case?

Question 3 What immediate action, if any, do you think the FDA should take?

Preparatory research

You find an excellent review article on the epidemiology of CJD (Brown, 1980). Through this and other sources (e.g., Benenson, 1995) you learn:

- CJD is a subacute degeneration of the central nervous system that occurs most often in people of late middle age. Death usually results within 6 months of onset.
- CJD affects men and women equally.
- Few cases have been reported in patients in their twenties and early thirties, although the youngest reported patient died at the age of 17. It is not until the age of 40 that CJD begins to occur with any consistency. Most cases occur in 50- to 75-year old people. The average age of death is approximately 60 years of age.
- In 1968, Gibbs and coworkers proved the CJD agent to be a membrane-associated unconventional virus of very small size and unusual resistance to physical and chemical means of inactivation. The pathogen is similar to the unconventional

agent that causes kuru in people and scrapie in sheep. (These agents are currently called ''prions.'')

- The virus can be found in cerebrospinal fluid and the central nervous system of patients. Transmission of the disease to primates is accomplished by inoculation through intracerebral and peripheral routes. The incubation period is up to 6 years when primates are inoculated through the intracerebral route.
- Incubations as long as 20 to 30 years are found in naturally occurring kuru.
- Approximately 10 000 children had received hGH from the National Hormone and Pituitary Program. (The program started in 1963.) There is no commercial distribution of hGH in the United States.
- A U.S. Vital Statistics report suggests there was a total of two deaths from CJD in persons less than 40 years of age in the United States in 1979. The age-specific mortality rate of the disease in persons less than 40 is therefore approximately 1 per 10 million.

Question 4 Knowing these new facts, would you have answered Question 1 any differently? If so, how?

How hGH is prepared

Every good epidemiologist can think of a reason why just about any association is biologically plausible (or, so it has been said). You learn that the hGH used during the interval when the case was exposed was extracted from human pituitaries obtained at autopsies. You also learn that: (a) it takes approximately 16 000 human pituitaries to make a single lot of hGH, (b) the average child undergoing treatment receives hGH injections 2–3 times a week for 4 yeas, and (c) a patient usually receives hGH from three different lots during each year of treatment.

Question 5 To how many pituitaries is the average hGH-treated child exposed?

Source at the agent

Although CJD is a rare disease among young people it is not as rare as you might suspect when deaths from all age groups are considered. For example, in 1979 there were 148 CJD cases among the 1 913 841 deaths in the United States.

Question 6 What proportion of deaths in the United States is CJD-related?

Question 7 If there were no screening criterion of pituitary donors, to how many infected pituitaries would the average treated child have been exposed?

Question 8 Even with careful screening of pituitaries, is it biologically plausible to suspect pituitaries as the vehicle of transmission? Why or why not?

More cases

Meanwhile, your medical officer friend has been on the phone to pediatric endocrinologists who have used hGH supplied by the National Hormone and Pituitary Program. She finds an additional two cases of CJD-like deaths occurring in the 10 000 hGM-exposed children. The first death is a 22-year old man treated with hGH from 1969 to 1977. The second death is a 34-year old man treated from 1963 to 1969. These two additional cases await pathological confirmation.

Question 9 How many cases of CJD were expected in the study population of 10 000 hGH-exposed children, assuming an expected rate of 1 in 10 million?

(*Hint*: Use Formula 19.3.) What is the SMR in this pituitary-exposed cohort? (SMR = a/μ where a represents the observed number of cases and μ represents the expected number of cases.)

Question 10 Assuming a random (Poisson) distribution of cases, and using the expected number of cases calculated in Question 9, calculate the Poisson probabilities of observing 0 cases in this cohort. What is the probability of observing exactly 1 case? What is the probability of observing exactly 2 cases? What is the probability of seeing 3 or more cases? Does this analysis confirm the existence of an epidemic? (*Hint*: Use formulas 19.1, 19.5, and 19.6.)

Question 11 What additional information would you like to know about the exposure of each individual case?

Question 12 What action, if any, do you think the FDA should take as a regulatory agency?

Question 13 The assistant secretary of health establishes an interagency task force to review the problem and make recommendations. You are assigned to the task force subgroup charged with directing future epidemiologic studies. Your subgroup recommends a case–control study to learn which lots of hGH were contaminated and a cohort study to learn the incidence of CJD overall and among subgroups. Each study will also explore host risk factors that determine the likelihood of disease. Discuss the basis of this recommendation.

Answers to case study: a drug–disease outbreak

Answer 1 The first step in the investigation is to verify the case's diagnosis. Assuming this case checks out, further investigation might be necessary.

Answer 2 Information concerning the ''who, what, when, where, why, and how'' of CJD and hGH should be studied. For example, the descriptive epidemiology of CJD must be researched (e.g., age and sex distribution of cases, age-specific rates, risk factors). In addition, we need to learn about how hGH is made and how it is used to treat dwarfism.

Answer 3 Although beginning immediate regulatory action is premature, active follow-up of people in the hGH-exposed cohort might be warranted.

Answer 4 Yes. Rarity of CJD in this age group, its transmissibility, and the inoculation of possible infected pituitary glands into susceptible individuals have heightened our level of suspicion that the association might be causal and that additional cases may be forthcoming.

Answer 5 Based on the stated assumptions, the average hGH-treated child would be exposed to 16 000 pituitaries/lot × 3 lots/year × 4 year = 192 000 pituitaries.

Answer 6 The proportion of deaths from CJD = 148/1 913 841 = 0.000 077.

Answer 7 Number of infected pituitary per hGH-treated child = 192 000 pituitaries × 0.000 077 = 14.8 infected pituitaries.

Answer 8 Yes. People dying of other causes may be incubating CJD.

Answer 9 Formula (19.2) states $\mu = (n)(\lambda_0)$, where μ represents the expected number of cases, n represents the size of the study population, and λ_0 represents the expected rate. Therefore, $\mu = (10\,000)(1/10\,000\,000) = 0.001$.
The SMR is equal to 3/0.001 = 3000.

Answer 10 We calculate the following Poisson probabilities by using Formula (19.1):

$$\Pr(X = 0) = \frac{\left(e^{-0.001}\right)\left(0.001^0\right)}{0!} = 0.999\,00$$

$$\Pr(X = 1) = \frac{\left(e^{-0.001}\right)\left(0.001^1\right)}{1!} = 0.001\,00$$

$$\Pr(X = 2) = \frac{\left(e^{-0.001}\right)\left(0.001^2\right)}{2!} = 0.000\,00$$

Thus, $\Pr(X \geq 3) = 1 - \Pr(X \leq 2) = 1 - (0.999\,000 + 0.000\,10 + 0.000\,00) = 0.000\,00$. Therefore, the epidemic is confirmed.

Answer 11 Knowing if a common ''hot'' lot was involved would be useful.

Answer 12 Recommendations for immediate actions are to cease use of hGH, notify all exposed individuals, and inform the public.

Answer 13 Case–control studies are well suited for the study of risk factors for rare diseases with long incubation (such as CJD). However, they are unable to estimate incidence. Therefore, a cohort study is necessary for this latter purpose. In addition, it may be noted that for these particular studies, we might expect difficulties in tracking exposures and identifying cases regardless of the study design. A problem of particular concern when considering the cohort design is the lengthy incubation period associated with the agent.

References—a drug–disease outbreak

Benenson, A.S. (ed.) (1995) *Control of Communicable Diseases in Man*, 16th edn, American Public Health Association, Washington, D.C.

Brown, P. (1980) An epidemiologic critique of Creutzfeldt–Jakob disease. *Epidemiologic Reviews*, **2**, 113–135.

Gibbs, C.J., Jr.,, Gajdusek, D.C., Asher, D.M., Alpers, M.P., Beck, E., Daniel, P.M., and Matthews, W.B. (1968) Creutzfeldt–Jakob disease (spongiform encephalopathy); *transmission to the chimpanzee*. *Science*, **161**, 388–389.

Chapter addendum 2 (case study)

Food borne outbreal in Rhynedale, California

Source: Adapted from DHEW/CDC training materials as presented in the University of California, Berkeley Infectious Disease Epidemiology course, Spring 1984.

Data: Data are available as rhynedal.rec (EpiInfo v6 format) and rhynedal.dbf on the *Epidemiology Kept Simple* website. (Data are also listed in Tables 12.3, 12.4, and 12.5.)

Comment

Food-borne disease outbreak investigations generally involve both a laboratory and an epidemiologic component. The **epidemiologic investigation** determines the number of cases, types and frequency of symptoms, location, date, and time of the

Table 12.3 Case status and demographic data: food-borne disease outbreak case study.

Variable names and descriptions:

REC	Record (identification) number	
CASE	Case of gastroenteritis (Y/N)	
SEX	Gender (M/F)	
AGE	Age (In years)	

REC	CASE	SEX	AGE	REC	CASE	SEX	AGE
1	Y	M	17	27	N	F	2
2	Y	F	8	28	N	F	15
3	Y	M	37	29	N	M	44
4	Y	F	7	30	N	F	42
5	Y	F	8	31	N	F	17
6	Y	M	26	32	N	F	19
7	Y	F	25	33	N	M	16
8	Y	F	5	34	N	M	3
9	Y	M	14	35	N	F	66
10	Y	M	48	36	N	F	18
11	Y	F	12	37	N	M	39
12	Y	M	10	38	N	F	17
13	Y	M	14	39	N	F	14
14	Y	M	13	40	N	M	46
15	Y	M	28	41	N	F	45
16	Y	F	9	42	N	F	16
17	Y	M	45	43	N	F	11
18	Y	F	35	44	N	M	26
19	Y	F	39	45	N	F	57
20	Y	M	10	46	N	M	60
21	Y	M	11	47	N	F	34
22	Y	M	7	48	N	F	16
23	Y	M	33	49	N	M	18
24	Y	F	41	50	N	F	4
25	Y	M	13	51	N	M	62
26	N	F	52				

suspect meal, onset time of symptoms in cases, history of food preparation and handling (including storage conditions), food and biologic samples (when possible), 24-hour food histories for all persons attending the suspect meal, and attack rates by food items eaten and not eaten. By piecing together these bits of information, the epidemiologists can usually determine the agent in question and its ultimate source.

The **laboratory investigation** is used to detect the presence of the agent in cases and the environment. The laboratory investigation is also useful in identifying other organisms in the environment that might be present in large enough numbers to suggest inadequacy of general sanitary conditions of food preparation. Laboratory procedures should include aerobic and anaerobic plate counts of samples collected during the epidemiologic investigation, with samples (e.g., stool, saliva, vomit) coming from both cases and foods, as appropriate, depending on the suspected etiologic agent. Culture and other diagnostic procedures aimed at the suspected agent should be pursued.

Table 12.4 Signs, symptoms, duration of illness, and incubation periods: food-borne disease outbreak case study.

Variable names and descriptions:

CRMP	Cramps (Y/N/·= missing)
DIAR	Diarrhea (Y/N/·= missing)
BDIAR	Bloody diarrhea (Y/N/·= missing)
NAUS	Nausea (Y/N/·= missing)
VOMIT	Vomiting (Y/N/·= missing)
FEV	Fever (Y/N/·= missing)
CHIL	Chills (Y/N/·= missing)
HEAD	Headache(Y/N/·= missing)
MYAL	Myalgia (Y/N/·= missing)
DUR	Duration of symptoms (in days)
INC	Incubation (hours between the beginning of supper and first symptoms)

Data:

REC	CRAMPS	DIAR	BDIAR	NAUS	VOMIT	FEV	CHIL	HEAD	MYAL	DUR	INC
1	Y	Y	N	Y	Y	Y	N	Y	Y	9	9
2	Y	Y	N	Y	Y	Y	Y	Y	Y	7	15
3	Y	Y	N	N	N	Y	Y	Y	N	3	17
4	Y	Y	·	Y	Y	N	Y	Y	·	7	19
5	Y	Y	N	N	N	Y	N	Y	Y	5	20
6	Y	Y	N	Y	Y	Y	Y	Y	Y	7	23
7	Y	Y	N	Y	N	N	N	Y	·	3	25
8	Y	Y	·	Y	Y	Y	Y	Y	·	7	25
9	Y	Y	N	Y	Y	Y	Y	Y	Y	8	26
10	Y	Y	N	N	N	Y	Y	Y	Y	2	18
11	Y	Y	N	Y	Y	Y	N	Y	Y	7	19
12	Y	Y	·	Y	N	Y	Y	Y	Y	8	32
13	Y	Y	N	N	N	Y	N	Y	Y	6	32
14	Y	Y	N	Y	Y	Y	Y	Y	Y	7	36
15	Y	Y	N	N	N	Y	Y	Y	Y	3	43
16	Y	Y	N	N	N	Y	Y	Y	Y	6	47
17	Y	Y	N	Y	Y	Y	Y	Y	Y	7	53
18	Y	Y	N	Y	Y	Y	Y	Y	Y	7	62
19	Y	Y	Y	Y	N	Y	Y	N	N	5	76
20	Y	Y	N	N	N	Y	N	Y	Y	4	77
21	Y	Y	N	Y	Y	Y	N	N	Y	7	89
22	Y	Y	·	N	N	Y	N	Y	·	7	97
23	Y	Y	N	Y	Y	Y	N	Y	Y	4	98
24	Y	Y	N	N	N	Y	Y	Y	Y	7	100
25	Y	Y	N	Y	N	Y	Y	Y	Y	8	111

Background

The local medical center in a rural California county notifies the county health department of a hospitalized case of gastroenteritis. From interviewing this patient, you discover that the case had attended a church supper in Rhynedale, California, seven days earlier. You find out that approximately 50 other individuals had attended this church supper and that several participants had become ill with symptoms similar to those seen in the sentinel case. Your assignment is to investigate this case and determine whether it was the result of food poisoning.

Table 12.5 Food histories: food-borne disease outbreak case study.

Variable names and descriptions:

SHRIMP	Shrimp salad (Y/N/·= missing)
OLIVE	Olives (Y/N/·= missing)
FRCHICK	Fried chicken (Y/N/·= missing)
BBQCHICK	Barbecued chicken (Y/N/·= missing)
BEANS	Beans (Y/N/·= missing)
POTSAL	Potato salad (Y/N/ — = missing)
GJEL	Green Jell-O (Y/N/ — = missing)
RJEL	Red Jell-O (Y/N/ — = missing)
MAC	Macaroni salad (Y/N/ — = missing)
RBEER	Root beer (Y/N/ — = missing)
ROLL	Rolls (Y/N/ — = missing)
BUTTER	Butter (Y/N/ — = missing)
DEVEGG	Deviled eggs (Y/N/ — = missing)
POTCHIP	Potato chips (Y/N/ — = missing)
PICK	Pickle (Y/N/ — = missing)
SCP	Strawberry cream pie (Y/N/ — = missing)
NCP	Neapolitan cream pie (Y/N/ — = missing)
CAKE	Cake (Y/N/ — = missing)
TOM	Tomato (Y/N/ — = missing)

Data: Part I (shrimp salad, olives, fried chicken, barbeque chicken, beans, potato salad, Green Jell-O, Red Jell-O, macaroni salad)

REC	CASE	SHRIMP	OLIVE	FRCHICK	BBQCHICK	BEANS	POTSAL	GJEL	RJEL	MAC
1	Y	N	Y	Y	Y	N	Y	Y	Y	N
2	Y	N	N	Y	—	N	N	Y	N	Y
3	Y	N	N	Y	Y	N	Y	—	N	N
4	Y	—	Y	Y	Y	N	Y	Y	Y	N
5	Y	N	Y	Y	—	Y	N	—	Y	Y
6	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
7	Y	—	Y	—	Y	—	Y	Y	N	N
8	Y	N	Y	Y	Y	Y	Y	N	Y	N
9	Y	N	Y	Y	Y	Y	Y	N	Y	N
10	Y	Y	Y	Y	N	N	Y	—	N	—
11	Y	Y	Y	Y	N	Y	Y	Y	N	N
12	Y	Y	Y	Y	N	Y	Y	N	Y	Y
13	Y	N	N	N	Y	N	N	N	N	N
14	Y	N	Y	Y	Y	N	Y	N	—	N
15	Y	Y	Y	Y	Y	Y	N	—	—	—
16	Y	N	N	Y	Y	N	N	—	N	Y
17	Y	N	Y	—	Y	Y	Y	—	N	Y
18	Y	Y	Y	Y	—	N	N	N	Y	N
19	Y	Y	N	Y	N	Y	Y	N	Y	N
20	Y	N	Y	Y	N	N	Y	N	—	Y
21	Y	N	—	—	Y	Y	N	Y	Y	N
22	Y	N	Y	—	Y	N	Y	N	Y	N
23	Y	Y	Y	Y	Y	Y	N	Y	—	Y
24	Y	N	Y	—	Y	Y	Y	Y	N	Y
25	Y	N	Y	Y	N	N	Y	N	Y	N
26	N	Y	N	N	N	Y	Y	—	N	Y
27	N	N	N	Y	N	N	Y	N	Y	—
28	N	N	Y	Y	N	.	Y	Y	—	Y

(*continued overleaf*)

Table 12.5 (*continued*)

REC	CASE	SHRIMP	OLIVE	FRCHICK	BBQCHICK	BEANS	POTSAL	GJEL	RJEL	MAC
29	N	N	N	Y	N	Y	Y	Y	N	N
30	N	N	Y	Y	N	N	N	Y	N	Y
31	N	N	N	Y	N	Y	Y	—	N	Y
32	N	Y	Y	Y	N	N	Y	N	Y	N
33	N	N	N	Y	N	N	Y	N	Y	Y
34	N	—	N	N	N	Y	Y	N	Y	Y
35	N	—	N	N	N	Y	Y	N	Y	Y
36	N	N	Y	Y	N	N	Y	—	—	Y
37	N	N	N	Y	N	N	Y	N	Y	N
38	N	N	N	Y	N	N	Y	—	—	N
39	N	N	N	Y	N	Y	Y	N	Y	N
40	N	N	N	Y	N	Y	Y	—	—	Y
41	N	N	N	Y	N	Y	N	N	Y	Y
42	N	N	Y	Y	N	Y	Y	Y	N	N
43	N	N	Y	Y	—	N	Y	Y	—	N
44	N	N	Y	Y	N	Y	Y	Y	N	Y
45	N	N	Y	Y	N	N	N	Y	.	N
46	N	N	N	Y	N	Y	Y	.	N	Y
47	N	N	Y	Y	N	Y	N	.	.	N
48	N	N	Y	Y	N	Y	N	Y	.	N
49	N	Y	Y	Y	N	.	Y	N	Y	Y
50	N	N	Y	Y	N	Y	Y	.	.	Y
51	N	Y	Y	Y	.	N	N	Y	.	Y

Data, *Part II* (*Root Beer*, *Rolls*, *Butter*, *Devilled Eggs*, *Potato Chips*, *Pickles*, *Strawberry Cream Ple*, *Neapolitan Cream Pie Cake*, *Tomato*)

REC	CASE	RBEER	ROLL	BUTTER	DEVEGG	POTCHIP	PICK	SCP	NCP	CAKE	TOM
1	Y	Y	N	N	.	Y	N	N	N	N	N
2	Y	Y	.	N	N	N	N	Y	N	N	N
3	Y	Y	N	N	Y	Y	Y	N	.	N	Y
4	Y	Y	N	N	Y	Y	N	N	.	N	N
5	Y	Y	N	N	Y	Y	N	N	Y	N	Y
6	Y	Y	N	N	N	Y	Y	Y	N	Y	Y
7	Y	Y	N	N	Y	Y	.	.	.	.	.
8	Y	Y	N	N	Y	Y	N	N	N	N	N
9	Y	Y	N	N	Y	Y	N	N	N	N	N
10	Y	Y	.	.	.	Y	Y	N	N	.	N
11	Y	Y	N	N	Y	Y	N	Y	N	N	N
12	Y	Y	N	N	Y	N	Y	N	N	.	N
13	Y	N	N	N	N	N	N	N	N	N	N
14	Y	Y	N	N	N	Y	Y	N	N	N	N
15	Y	Y	N	N	Y	Y	Y	Y	N	N	.
16	Y	Y	Y	Y	N	N	Y	Y	N	N	N
17	Y	Y	Y	Y	N	N	Y	Y	N	N	N
18	Y	Y	N	N	Y	Y	Y	Y	N	N	N
19	Y	Y	N	N	N	Y	N	N	N	N	N
20	Y	Y	N	N	Y	Y	N	N	N	N	N
21	Y	N	.	.	N	Y	Y	Y	N	.	Y
22	Y	Y	N	N	N	Y	Y	Y	N	N	N
23	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N
24	Y	Y	Y	Y	N	Y	N	N	N	.	N

(*continued overleaf*)

Table 12.5 (*continued*)

REC	CASE	RBEER	ROLL	BUTTER	DEVEGG	POTCHIP	PICK	SCP	NCP	CAKE	TOM
25	Y	Y	N	N	N	Y	N	N	N	N	Y
26	N	Y	N	N	N	.	N	N	N	N	Y
27	N	Y	N	N	N	Y	N	N	N	N	Y
28	N	Y	Y	Y	Y	Y	N	N	N	Y	Y
29	N	Y	N	N	.	Y	Y	N	N	N	N
30	N	Y	Y	Y	N	Y	Y	N	N	N	Y
31	N	Y	N	N	N	Y	Y	N	N	Y	Y
32	N	Y	Y	Y	N	Y	N	Y	N	N	N
33	N	Y	N	N	N	Y	N	N	N	N	N
34	N	Y	N	N	N	Y	N	N	N	N	Y
35	N	Y	N	N	N	Y	N	N	N	N	Y
36	N	—	Y	—	N	Y	N	—	—	N	Y
37	N	Y	N	N	N	Y	N	N	N	—	Y
38	N	Y	N	N	Y	Y	N	N	N	—	N
39	N	N	Y	Y	N	N	N	.	N	N	Y
40	N	Y	Y	Y	N	N	N	N	Y	N	Y
41	N	Y	Y	Y	N	N	Y	N	N	N	N
42	N	Y	N	N	N	N	N	N	N	N	N
43	N	Y	Y	Y	—	Y	N	N	N	N	N
44	N	Y	N	N	Y	N	Y	N	N	Y	N
45	N	Y	Y	Y	Y	N	N	N	Y	N	Y
46	N	Y	Y	Y	Y	N	Y	N	—	N	Y
47	N	Y	N	N	Y	N	Y	Y	N	N	N
48	N	Y	N	N	N	N	N	N	N	N	Y
49	N	Y	Y	—	N	N	N	Y	N	N	Y
50	N	Y	N	—	N	Y	Y	—	—	Y	N
51	N	N	Y	Y	Y	Y	Y	N	N	—	Y

The Locale

Rhynedale is a small, unincorporated town of 581 residents in northern California, not far from Sacramento.

Question 1 Discuss how you might prepare for the field work.

The church supper

Between the hours of 6:00 p.m. and 9:30 p.m. on Saturday, 23 July, a Rhynedale community church held a pot luck supper in a wooded area near its church. The attending families each brought food from home, which was laid out on tables for all to share. Though many brought similar foods, none of these foods were mixed, and all food remained in its original container. No caterers or other persons were involved in handling the food. The bulk of the food was eaten between 6:30 and 7:30 p.m. All cases that were contacted denied having a diarrheal illness before or during the supper and denied knowledge of the same for others attending the supper.

The involved group

Fifty-one people from 16 family groups attended this affair. Thirteen of the families were from the Rhynedale area. One family was from a nearby town. Two families were from out-of-state.

Materials and methods

You place phone calls to all families who had attended the supper. Family rosters are recorded and, for anyone reporting symptoms, the course and duration of their illness are described by onset, duration, symptoms, and the need for physician care or hospitalization. Food histories are obtained on all supper attendees—whether ill or not—for each item served at the supper. Inquiries are made of pet and other animal contacts, and of prior contacts with persons known to have had diarrheal illnesses.

You request each family to furnish a list of foods they brought to the supper. You also inquire where they purchased these foods, how they were prepared, what foods they took home from the supper, and whether food items were still available for laboratory testing. Visits are made to two families who did not have telephones. (This is a fairly old case study.) Out-of-state parties are not reached, although multiple telephone calls are attempted.

Data

Each study subject is uniquely identified with a record (REC) number. Information on case status and demographic factors appear in Table 12.3. Information on signs, symptoms, duration of illness, and time of onset relative to the beginning of the church supper appears in Table 12.4. Note that a ''—'' indicates missing data. Information about food items consumed by people attending the church supper appears in Table 12.5.

Question 2 Draw an epidemic curve. Determine the minimum, maximum, and median incubation times.

The median is the data point that is greater than or equal to half the values in the sample. Before calculating the median, data are sorted in ascending order and ranked from 1 to n, where n represents the number of cases. The median value is the value halfway down the rank-ordered list. If n is odd, this corresponds to the value of the observation with rank $(n + 1)/2$. If n is even, this corresponds to the average of values associated with ranks $n/2$ and $(n/2) + 1$.

Question 3 Determine the frequency of symptoms by filling in Table 12.6. Exclude people with missing values from numerators and denominators of the frequency calculation in question.

Question 4 Calculate the food-specific attack rates (incidence proportions) and risk ratios associated with each food item. Attack rates are calculated as follows:

$$\text{Attack rate}_{\text{exposed}} = \frac{\text{no. of people who ate food and became ill}}{\text{total no. of people who ate food}} \tag{12.1}$$

$$\text{Attack rate}_{\text{unexposed}} = \frac{\text{no. of people who did not eat food and became ill}}{\text{total no. of people who did not eat food}} \tag{12.2}$$

The risk ratio (RR) is

$$\text{RR} = \frac{\text{attack rate}_{\text{exposed}}}{\text{attack rate}_{\text{unexposed}}} \tag{12.3}$$

Write your answers in Table 12.7. Based on your calculations, what is the most likely source of exposure to the pathogen?

Table 12.6 Frequency of symptoms: food-borne disease outbreak case study.

Symptom	Number reporting symptom	Number of cases responding to question	Percentage
Cramps			
Diarrhea			
Bloody diarrhea			
Nausea			
Vomiting			
Fever			
Chills			
Headache			
Myalgia			

Table 12.7 Attack rates and relative risks: food-borne disease outbreak case study.

Food exposure	Attack rate in consumers	Attack rate in nonconsumers	Risk ratio
Shrimp salad	8/12 (66.7%)	15/35 (42.9%)	1.56
Olives			
Fried chicken			
Barbecued chicken			
Beans			
Potato salad			
Green Jell-O			
Red Jell-O			
Macaroni salad			
Root beer			
Rolls			
Butter			
Deviled eggs			
Potato chips			
Pickle			
Strawberry cream pie			
Neopolitan cream pie			
Cake			
Tomato			

Question 5 Use the Abbreviated Compendium of Acute Foodborne Gastrointestinal Diseases that appears as Table 12.8 to create a list of the most likely agent. Base this list on the typical signs, incubation period, and most likely food source of the pathogen.

Question 6 What are the mechanisms by which such vehicles usually become contaminated with the pathogen you suspect caused this outbreak?

Question 7 What measures are possible to prevent such contamination?

Question 8 What can be done to prevent illness if foods do become contaminated?

Table 12.8 Abbreviated compendium of acute food-borne gastrointestinal diseases.

Agent	Incubation period usual (and range)	Symptoms[a] (partial list)	Pathophysiology	Characteristic foods	Specimens[a]
I. Diseases typified by vomiting after a short incubation period with little or no fever					
A. *Staphytoccocus aureus*	2–4h (1–6 h)	N, C, V; D, F may be present	Preformed enterotoxin	Sliced/chopped ham and meats, custards, cream fillings	*Food*: enterotoxin assay (FDA), culture for quantitation and phage typing of staph, gram stain *Handlers*: culture nares, skin, skin lesions, and phage-type staph *Cases*: culture stool and vomitus, phage-type staph
B. *Bacillus cereus*	2–4h (1–6 h)	N, V, D	? Preformed enterotoxin	Fried rice	*Food*: culture for quantitation *Cases*: stool culture
C. Heavy metals 1. Cadmium 2. Copper 3. Tin 4. Zinc	5–15 min (1–60 min)	N, V, C, D		Foods and beverages prepared/stored/cooked in containers coated/lined/contaminated with offending metal	Toxicologic analysis of food container, vomitus, stomach contents, urine, blood, faces
II. Diseases typified by diarrhea after a moderate to long incubation period, often with fever					
A. *Clostridium perfringens*	12 h (8–16 h)	C, D (V, F rare)	Enterotoxin formed *in vivo*	Meat, poultry	*Food*: enterotoxin assay done as research procedure by FDA, culture for quantitation and serotyping

					Cases: culture feces for quantitation and serotyping of *C. perfringens*; test for enterotoxin in stool *Cases*: culture feces for quantitation and serotyping of *C. perfringens*
B. *Salmonella* (nontyphoid)	12–36 h (6–72 h)	D, C, F, V, H septicemia or enteric fever	Tissue invasion	Poultry, eggs, meat, raw milk (cross-contamination important)	*Food*: culture with serotyping *Cases*: stool culture with serotyping *Handlers*: stool culture with serotyping as a secondary consideration
C. *Vibrio parahaemolyticus*	12 h (2–48 h)	C, D N, V, F, H, B	Tissue invasion, and possible enterotoxin	Seafood	*Food*: culture on TCBS, serotype, Kanagawa test *Cases*: stool cultures on TCBS, serotype, Kanagawa test
D. *Escherichia coli enterotoxigenic*	16–48 h	D, C	Enterotoxin	Uncooked vegetables, salads, water, cheese	*Food*: culture and serotype *Cases*: stool cultures; serotype and enterotoxin production, invasiveness assay

(*continued overleaf*)

Table 12.8 (*continued*)

Agent	Incubation period usual (and range)	Symptoms[a] (partial list)	Pathophysiology	Characteristic foods	Specimens[a]
E. coli enteroinvasive	16–48 h	C, D, F, H	Tissue invasion	Same	*Cases*: stool cultures; serotype and enterotoxin production. Look for common serotype in food and cases not found in controls; DNA probes
E. coli enterohemorrhagic (*E. coli* O157.H7 and others)	48–96 h	B, C, D, H; F infrequent	Cytotoxin	Beef, raw milk, water	Stool cultures on MacConkey's sorbitol; serotype
E. *Bacillus cereus*	8–16 h	C, D	? Enterotoxin	Custards, cereals, puddings, sauces, meat loaf	*Food*: culture *Cases*: stool cultures
F. *Shigella*	24–48 h	C, F, D, B, H, N, V	Tissue invasion	Foods contaminated by infected foodhandler; usually not food-borne	*Food*: culture and serotype *Cases*: stool culture and serotype *Handlers*: stool culture and serotype
G. *Yersinia enterocolitica*	3–5 days (usual) range unclear	F, D, C, V, H	Tissue invasion, and possible enterotoxin	Pork products and foods contaminated by infected human or animal	*Food*: culture *Cases*: stool, blood cultures, serology *Handlers*: stool cultures

H. *Vibrio cholerae* O-1	24–72 h	D, V	Enterotoxin formed *in vivo*	Shellfish, water or foods contaminated by infected person or obtained from contaminated environmental source	*Food*: culture on TCBS, serotype *Cases*: stool cultures on TCBS, serotype Send all isolates to CDC for confirmation and toxin assay
I. *Vibrio cholerae* non O-1	16–72 h	D, V	Enterotoxin formed *in vivo* ? Tissue invasion	Shellfish	*Food*: culture on TCBS, serotype *Cases*: stool cultures on TCBS, serotpye
J. *Campylobacter jejuni*	3–5 days	C, D, B, F	Unknown	Raw milk, poultry, water	*Food*: culture on selective media (5%O_2, 42 °C) *Cases*: culture on selective media (5%O_2, 42°C) serology
K. Parvovirus-like agents (Norwalk, Hawaii, Colorado, cockle agents)	16–48 h	N, V, C, D	Unknown	Shellfish, water	Stool for immune electron microscopy and serology by special arrangement
L. Rotavirus	16–48 h	N, V, C, D	Unknown	Food-borne transmission not well documented	*Cases*: stool examination by EM or ELISA; serology

(*continued overleaf*)

Table 12.8 (*continued*)

Agent	Incubation period usual (and range)	Symptoms[a] (partial list)	Pathophysiology	Characteristic foods	Specimens[a]
III. Botulism					
Clostridium botulinum	12–72 h	V, D, descending paralysis	Performed toxin	Improperly canned or similarly preserved foods	*Food*: toxin assay *Cases*: serum and feces for toxin assay by CDC or state lab; stool culture for *C. botulinum*
IV. Diseases most readily diagnosed from the history of eating a particular type of food					
A. Poisonous mushrooms	Variable	Variable		Wild mushrooms	*Food*: speciation by mycetologist.
B. Other poisonous plants	Variable	Variable		Wild plant	*Cases*: vomitus, blood, urine *Food*: speciation by botanist; feces may sometimes by helpful in confirmation
C. Scombroid fish poisoning		N, C, D, H, flushing, urticaria	Histamine	Mishandled fish (i.e., tuna)	*Food*: Histamine levels
Ciguatera poisoning		D, N, V, paresthesias, reversal of temperature sensation	Ciguatoxin	Large ocean fish (i.e., barracuda, snapper)	*Food*: Stick test for ciguatoxin (not widely available)
D. Other poisonous sources	Variable	Variable	Variable		

[a] Abbreviations: B = bloody stools, C = cramps, D = diarrhea, F = fever, H = headache, N = nausea, V = vomiting.
Source: Centers for Disease Control. (1992). ''*Oswego*:'' *An Outbreak of Gastrointestinal Illness Following a Church Supper*. Association of Teachers of Preventive Medicine, Washington, D.C.

Table 12.9 Frequency of symptoms: food-borne disease outbreak case study answer key.

Symptom	Number reporting symptom	Number responding to question	Percentage (%)
Cramps	25	25	100
Diarrhea	25	25	100
Bloody diarrhea	1	21	5
Nausea	17	25	68
Vomiting	12	25	48
Fever	23	25	92
Chills	16	25	64
Headache	23	25	92
Myalgia	19	21	90

Table 12.10 Attack rates and relative risks: food-borne disease outbreak case study answer key.

Food exposure	Attack rate in consumers	Attack rate in nonconsumers	Risk ratio
Shrimp salad	8/12 (67%)	15/35 (43%)	1.6
Olives	19/32 (59%)	5/18 (28%)	2.1
Fried chicken	19/42 (45%)	1/4 (25%)	1.8
Barbecued chicken	16/16 (100%)	6/30 (20%)	5.0
Beans	12/26 (46%)	12/22 (55%)	0.8
Potato salad	17/37 (46%)	8/14 (57%)	0.8
Green Jell-O	8/17 (47%)	11/20 (55%)	0.9
Red Jell-O	12/21 (57%)	9/16 (56%)	1.0
Macaroni salad	9/24 (38%)	14/24 (58%)	0.6
Root beer	23/46 (50%)	2/4 (50%)	1.0
Rolls	4/16 (25%)	18/32 (56%)	0.4
Butter	4/14 (29%)	19/32 (59%)	0.5
Deviled eggs	12/19 (57%)	11/28 (39%)	1.6
Potato chips	20/35 (57%)	5/15 (33%)	1.7
Pickles	12/21 (76%)	12/29 (41%)	1.4
Strawberry cream pie	10/13 (33%)	14/34 (41%)	1.9
Neapolitan cream pie	1/3 (20%)	21/42 (50%)	0.7
Cake	1/5 (24%)	19/38 (50%)	0.4
Tomato	5/21	18/28 (64%)	0.4

Answers to case study: food-borne disease outbreak

Answer 1 The following text comes from the CDC *Principles of Epidemiology: Self-Study Course* (1992a):

> Anyone about to embark on an outbreak investigation should be well prepared before leaving for the field. Preparations can be grouped into three categories: investigation, administration, and consultation. Good preparation in all three categories will facilitate a smooth field experience.
>
> *Investigation* First, as a field investigator, you must have the appropriate scientific knowledge, supplies, and equipment to carry out the investigation. You should discuss the situation with someone knowledgeable about the disease and about field investigations, and review the applicable literature. You should assemble useful references such as journal articles and sample questionnaires. Before leaving for a field investigation, consult laboratory staff to ensure that you take the proper laboratory material and know the proper collection, storage, and transportation techniques. Arrange for a portable computer, dictaphone, camera, and other supplies.
>
> *Administration* Second, as an investigator, you must pay attention to administrative procedures. In a health agency, you must make travel and other arrangements and get them approved. You may also need to take care of personal matters before you leave, especially if the investigation is likely to be lengthy.
>
> *Consultation* Third, as an investigator, you must know your expected role in the field. Before departure, all parties should agree on your role, particularly if you are coming from ''outside'' the local area. For example, are you expected to lead the investigation, provide consultation to the local staff who will conduct the investigation, or simply lend a hand to the local staff? In addition, you should know who your local contacts will be. Before leaving, you should know when and where you are to meet with local officials and contacts when you arrive in the field. (pp. 353–354)

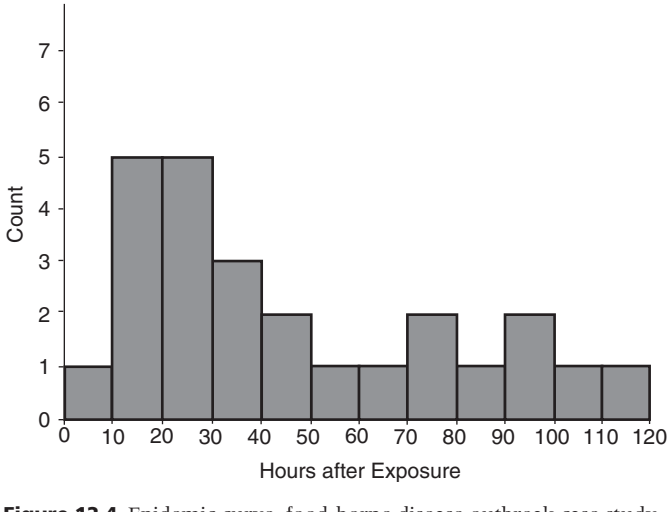


Figure 12.4 Epidemic curve, food-borne disease outbreak case study.

Answer 2 The epidemic curve appears as Figure 12.4. This curve is unimodal with a long right tail, reflecting the point-source nature of this epidemic. The median incubation time is 32 h. The minimum incubation time is 9 h, and the maximum is 111 h.

Answer 3 Frequencies of symptoms are seen in Table 12.9. The disease is characterized by cramping, diarrhea, nausea, vomiting, fever, chills, headache, and myalgia. Bloody diarrhea is generally absent.

Answer 4 Attack rates and risk ratios associated with each food are listed in Table 12.10. Based on these data, the most likely source of the pathogen is the barbecued chicken (RR = 5.0).

Answer 5 The most likely agent is *Salmonella*. *Salmonella* is the only agent typified by diarrhea, a moderate-to-long incubation period, cramps, fever, and vomiting. It is characteristically transmitted in poultry. The second most likely agent is *Shigella*, although *Shigella* is often characterized by the presence of blood diarrhea, which was absent from this outbreak.

Answer 6 Salmonella is often endemic in chicken populations. Contamination may spread during food handling, either at the butcher or at home.

Answer 7 Thorough cooking of all foodstuffs derived from animal sources, especially from fowl, can prevent such contamination. Also, cross-contamination with raw poultry must be avoided after cooking is completed. Food handlers should be educated as to the necessity of refrigeration, food preparation, and maintenance of a sanitary environment.

Answer 8 Proper cooking and heating of foodstuff will prevent illness.

CHAPTER 13
Confidence Intervals and *p*-Values

13.1 Introduction
- Parameters and estimates
- Population and sample
- Statistical inference

13.2 Confidence intervals
- Estimation
- Confidence intervals for proportions (incidence proportion and prevalence)
  - Small sample methods
- Confidence intervals for rates
- Confidence intervals for proportion ratios (risk ratios and prevalence ratios)
- Confidence intervals for rate ratios
- Confidence intervals for proportion differences (risk differences and prevalence differences)
- Confidence intervals for rate differences
- Confidence intervals for odds ratios, independent samples
- Confidence intervals for odds ratios, matched pairs

13.3 *p*-Values
- Hypothesis tests of statistical significance
  - Null and alternative hypotheses
  - Basis of the *p*-value
- Fallacies of *p*-values and statistical testing
- Testing a proportion
- Testing a rate
- Chi-square test of association
- Fisher's exact test
- Testing independent rates
- McNemar's test for matched pairs

13.4 Minimum Bayes factors
- Introduction
- Bayes factor
- Interpretation of the Bayes factor
- Prior odds
- Method to calculate a minimum Bayes factor

References

> Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise.
>
> *John W. Tukey (1962, pp. 13–14)*

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

13.1 Introduction

Parameters and estimates

This chapter discusses the traditional methods used to address random error in quantitative research. Before discussing these methods, we should recall the distinction between statistical parameters and estimates as discussed in Chapter 9. The term **parameter** refers to an error-free numerical constant that statistically describes a numerical characteristic of a population. We cannot in fact know the value of the parameter exactly but can estimate it statistically. However, the calculated statistical **estimate** will be imperfect due to random and systematic error. Methods presented in this chapter address the problem of random error in statistical estimates but have no influence on the problem of systematic error.

Let us introduce new **notation** in this chapter that preserves the distinction between parameters and estimates by using Greek letters to denote parameters, while using the same Greek letter with an overhead hat ($\hat{}$)] to denote analogous estimators. For example, ϕ is used to denote a risk ratio parameter, while $\hat{\phi}$ is used to denote a risk ratio estimator. The only exception to the convention of using Greek characters for parameter is seen in using the Roman letter p to denote the binomial proportion parameter and the symbol $\hat{p}$ to denote its estimator. This exception is adopted to avoid confusion with the constant π and maintain conventions of accepted biostatistical usage.

Population and sample

Understanding the population that is the source of the parameter we want to learn about is fundamental to statistical inference. This topic was introduced in Section 9.2 and bears further comment here. Statistical populations may be real or hypothetical, depending on the object of the research. **Real populations** are composed of a finite number of potential observations, as describes the situation when conducting a prevalence survey. In contrast, **hypothetical populations** represent an infinite number of potential observations from which to sample. While imagining a real population is relatively undemanding, conceiving of a hypothetical population is not always self-evident. The difference between a real population and hypothetical population can be made more tangible by considering the Framingham Heart Study. While it is true that the Framingham cohort is real and exists in eastern Massachusetts (United States), most inferences based on these data are used to learn about general physiologic relations between various risk factors (e.g., hypercholesterolemia) and heart disease. Since these generalizations go beyond the present Framingham population, data are seen as representing a sample from a hypothetical ''superpopulation'' of general causal phenomena.

Statistical inference

Regardless of whether a study is based in a real population or hypothetical population, the two traditional methods of inferring statistical parameters are estimation and hypothesis (significance) testing. Examples will introduce their use. It is common for

epidemiologists to want to learn about the prevalence of a condition (e.g., smoking) in a population based on the prevalence of the condition in a sample. In a given study, the inference may be ''25% of the population smokes'' (**point estimation**). In addition, estimation takes the form of an interval, such as ''we are 95% confident that between 20% and 30% of the population smokes'' (**interval estimation**). Finally, the epidemiologist might want to test the claim that the prevalence of smoking has changed. Under such instances, a categorical ''yes'' or ''no'' conclusion is sought (**hypothesis testing**). Our first order of business is to present methods of calculating confidence intervals.

13.2 Confidence intervals

Estimation

Estimation—the process of using data to ''locate'' population parameters—comes in two forms: point estimation and interval estimation. For example, when in the past we had calculated a risk difference of, say, 10%, this was the point estimate for risk difference parameter δ. **Interval estimation** surrounds the point estimate with a margin of error, thus creating a **confidence interval** (Figure 13.1). For example, a 95% confidence interval for a risk difference might be $0.10 \pm .02$. This is written (0.08, 0.12), where 0.08 is the **lower confidence limit** (LCL) of the interval and 0.12 is the **upper confidence limit** (UCL). The width of this confidence interval—simply its upper limit minus its lower limit—is a measure of the estimate's precision. (For risk ratios and rate ratios, the ratio of the upper and lower confidence limits is a comparable measure of precision.) Wide confidence intervals indicate low precision, and narrow confidence intervals indicate high precision. Large studies tend to derive narrow confidence intervals (precise estimates). Small studies tend to derive wide confidence intervals (imprecise estimates). Formulas and illustrations follow.

Confidence intervals for proportions (incidence proportion and prevalence)

Both incidence proportions (risks) and prevalences are mathematical proportions. Assuming observations are independent, the number of cases in a given sample will follow a binomial distribution with parameters n and p, where n represents the sample size and p represents the incidence proportion or prevalence parameter. The estimator

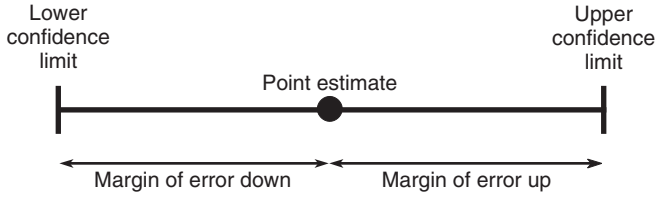


Figure 13.1 Representation of a confidence interval.

of this parameter, denoted $\hat{p}$ (''p hat''), is simply the sample proportion:

$$\hat{p} = \frac{A}{n} \tag{13.1}$$

where A represents the number of cases in the sample and n represents the sample size. For example, a sample of 57 people in which 17 smoke demonstrates $\hat{p} = 15/57 = 0.298$ for the characteristic of smoking.

Let $\hat{q} = 1 - \hat{p}$. When $n\hat{p}\hat{q} \geq 10$, a **normal approximation** to the binomial can be used to calculate a confidence 95% interval for p as follows:

$$\hat{p} \pm (1.96)(\mathrm{SE}_{\hat{p}}) \tag{13.2}$$

where $\mathrm{SE}_{\hat{p}}$ represents the estimated **standard error of the proportion**:

$$\mathrm{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Other levels of confidence for this and other confidence intervals in this chapter based on normal error distributions are calculated by replacing the 1.96 in formulas with 1.645 for 90% confidence intervals and 2.58 for 99% confidence intervals. For example, a 90% confidence interval for p is given by $\hat{p} \pm (1.645)(\mathrm{SE}_{\hat{p}})$. A 99% confidence interval for p is given by $\hat{p} \pm (2.58)(\mathrm{SE}_{\hat{p}})$.

Illustrative Example 13.1 Confidence interval for a proportion (large sample method)

For the random sample of $n = 57$ with $A = 17$, $\hat{p} = 17/57 = 0.2982$, $\hat{q} = 1 - 0.2982 = 0.7018$, and $n\hat{p}\hat{q} = (57)(0.2982)(1 - 0.2982) = 11.9$. Therefore, the normal approximation to the binomial (large sample method) can be applied. The estimated standard error of the proportion is $\mathrm{SE}_{\hat{p}} = \sqrt{\frac{(0.2982)(1 - 0.2982)}{57}} = 0.0606$ and a 95% confidence interval for $p = 0.2982 \pm (1.96)(.0606) = 0.2982 \pm 0.1188 = (0.179, 0.417)$.

Small sample methods

In small samples the above method should be voided in preference for either a quadratic method (Formula 13.3) or exact binomial method. The lower confidence limit (p_{LCL}) and upper confidence limit (p_{UCL}) for a 95% confidence interval for p by the **quadratic method** (Fleiss, 1981, Section 1.4) are

$$\hat{p}_{\mathrm{LCL}} = \frac{\left(2n\hat{p} + 1.96^2 - 1\right) - 1.96\sqrt{1.96^2 - (2 + 1/n) + 4\hat{p}\,(n\hat{q} + 1)}}{2\left(n + 1.96^2\right)}$$

$$\hat{p}_{\mathrm{UCL}} = \frac{\left(2n\hat{p} + 1.96^2 \quad + \quad 1\right) - 1.96\sqrt{1.96^2 + (2 + 1/n) + 4\hat{p}\,(n\hat{q} + 1)}}{2\left(n + 1.96^2\right)} \tag{13.3}$$

Coverage of the method used to calculate **exact binomial** confidence limits is beyond the scope of this book. Fortunately, there exist reliable public domain computer programs that use this method to compute confidence limits. The two that have been emphasized in this book are **OpenEpi.com** (Dean *et al.*, 2006) and **WinPEPI** (Abramson, 2011). For example, OpenEpi → Counts → Proportions derives a mid-P exact 95% confidence interval for p of 0.191–0.426.

Confidence intervals for rates

When estimating an incidence rate based on a person-time denominator, the number of cases in a given sample is assumed to follow a Poisson distribution with expectation μ. Let A represent the observed number of cases in a sample comprising T person-years of population experience. The point estimator of rate parameter λ is

$$\hat{\lambda} = \frac{A}{T} \tag{13.4}$$

A Fisher's exact **95% confidence interval for rate parameter** λ is calculated in two steps. First, the lower confidence limit (A_{LCL}) and the upper confidence limit (A_{UCL}) for the expected number of cases are determined using the Poisson limits in Appendix 1. Then, these confidence limits are expressed relative to the person-time (T) in the sample:

$$\left(\frac{A_{\text{LCL}}}{T}, \ \frac{A_{\text{UCL}}}{T}\right) \tag{13.5}$$

Illustrative Example 13.2 Confidence interval for a rate

Suppose 25 deaths are found in 4054 person-years of observation. The mortality rate,

$$\hat{\lambda} = \frac{25 \text{ people}}{4054 \text{ person-years}} = 0.0062 \text{ year}^{-1}$$

This may be expressed with a population base of 1000 as 6.2 per 1000 person-years. A 95% confidence interval for the rate is calculated in two steps. The confidence limits for the number of cases is (16.18, 36.90)—see Appendix 1—and a 95% confidence for λ is (16.18/4054 person-years, 36.90/4054 person-years) = (0.0040, 0.0091) year^{-1}. Expressed with a population base of 1000, the 95% confidence interval for λ is (4.0, 9.1) per 1000 person-years. The same results can be derived with *OpenEpi.com* → Person-time → 1 Rate → Fisher's exact.

Confidence intervals for proportion ratios (risk ratios and prevalence ratios)

The ratio of two incidence proportions is a risk ratio. The ratio of two prevalences is a prevalence ratio. Let ϕ represent a **proportion ratio parameter** (either a risk ratio or prevalence ratio). The point estimator of the proportion ratio parameter is

$$\hat{\phi} = \frac{\hat{p}_1}{\hat{p}_0} \tag{13.6}$$

where $\hat{p}_1$ represents the proportion in the exposed group and $\hat{p}_0$ represents the proportion in the nonexposed group. Table 13.1 displays additional notation. The sampling distribution of the *natural logarithm* of $\hat{\phi}$ is approximately normal. Therefore, we transform the risk ratio estimate into a natural logarithmic (ln) scale before calculating its confidence interval. (Use the ''ln'' key on your calculator.)

The **standard error of the natural log of the proportion ratio** is

$$\text{SE}_{\ln \hat{\phi}} = \sqrt{\frac{1 - \hat{p}_1}{N_1 \hat{p}_1} + \frac{1 - \hat{p}_0}{N_0 \hat{p}_0}} \tag{13.7}$$

Table 13.1 Notation for incidence proportion, prevalence, and case–control data.

	Disease +	Disease −	
Exposure +	A_1	B_1	N_1
Exposure −	A_0	B_0	N_0
	M_1	M_0	N

The **95% confidence interval for ln** ϕ is given by

$$\ln \hat{\phi} \pm (1.96)\left(\mathrm{SE}_{\ln \hat{\phi}}\right) \tag{13.8}$$

Antilogs (exponents) of the confidence limits are taken to convert them into a nonlogarithmic scale. (The antilog key may be labeled e on your calculator or might be accessed by pressing ''inverse ln.'')

Illustrative Example 13.3 Confidence interval for a risk ratio

The drug cytarabine is used for bone marrow ablation in preparation for transplantation. Even under the best of circumstances, this drug is associated with a high risk of cerebellar toxicity. There was a suspicion that the drug produced by a generic manufacturer presented a greater risk than the innovator product to those who used it. Table 13.2 contains data from a study that addressed this question. Based on these data, the risk of toxicity with the generic drug ($\hat{p}_1$) = 11/25 = 0.440. The risk of cerebellar toxicity with the innovator drug ($\hat{p}_0$) = 3/34 = 0.088. The risk ratio estimate $\hat{\phi}$ = 4.99 and $\ln \hat{\phi}$ = ln(4.99) = 1.607. The standard error estimate is

$$\mathrm{SE}_{\ln \hat{\phi}} = \sqrt{\frac{1 - 0.440}{(25)\,(0.440)} + \frac{1 - 0.088}{(34)\,(0.088)}} = 0.5964$$

and a 95% confidence interval for $\ln \phi = 1.607 \pm (1.96)(0.5964) = (0.438, 2.776)$. The 95% confidence limits for our risk ratio = $e^{0.438, 2.776}$ = (1.55, 16.05). An identical result is derived with *OpenEpi.com* → Counts → Two-by-Two Table.

Table 13.2 Data for Illustrative Examples 13.3. Cerebellar toxicity according to use of generic or innovator cytarabine, cohort study.

	Toxicity +	Toxicity −	
Generic +	11	14	25
Generic −	3	31	34
	14	45	59

$$\hat{p}_1 = \frac{11}{25} = 0.4400$$

$$\hat{p}_0 = \frac{3}{34} = 0.0882$$

Source: Jolson *et al*. (1992).

Confidence intervals for rate ratios

Let ω represent the rate ratio parameter. The rate ratio estimator is

$$\hat{\omega} = \frac{\hat{\lambda}_1}{\hat{\lambda}_0} \tag{13.9}$$

where $\hat{\lambda}_1$ represents the rate in the exposed group (= A_1/T_1) and $\hat{\lambda}_0$ represents the rate in the nonexposed group (= A_0/T_0). With moderate to large samples, the random error distribution of the natural log of the estimator ($\ln \hat{\omega}$) is approximately normal with a **standard error** of

$$\mathrm{SE}_{\ln \hat{\omega}} = \sqrt{\frac{1}{A_1} + \frac{1}{A_0}} \tag{13.10}$$

The **95% confidence interval for ln ω** is calculated in the usual manner:

$$\ln \hat{\omega} \pm (1.96)\left(\mathrm{SE}_{\ln \hat{\omega}}\right) \tag{13.11}$$

The **95% confidence limits for ω** is derived by taking the exponents of these limits.

Illustrative Example 13.4 Confidence interval for a rate ratio

Table 13.3 presents data from a study on physical fitness and mortality (Blair *et al*., 1995). This study found 25 deaths in 4054 person-years in men who went from the physically unfit to the physically fit category ($\hat{\lambda}_1$ = 25/4054 person-years = 0.0062 year^{-1}). It found 32 deaths in 2937 person-years in the men who remained in the unfit category ($\hat{\lambda}_0$ = 32/2937 person-years = 0.0109 year^{-1}). Thus, $\hat{\omega} = \dfrac{0.0062 \text{ year}^{-1}}{0.0109 \text{ year}^{-1}} = 0.57$ and $\ln(\hat{\omega}) = \ln(0.57) = -0.562$. The standard error on a logarithmic scale (base *e*) is $\mathrm{SE}_{\ln \hat{\omega}} = \sqrt{\dfrac{1}{25} + \dfrac{1}{32}} = 0.2669$ and a 95% confidence interval for $\ln \omega = -0.562 \pm (1.96)(0.2669) = -0.562 \pm 0.523 = (-1.085, -0.039)$. The 95% confidence limits for ω is $e^{(-1.085, -0.039)} = (0.34, 0.96)$.

We can use *OpenEpi.com* → Person-time → Compare 2 rates to derive exact confidence intervals for rate ratios. WinPEPI → Compare2 → D. Rates with person-time denominators can also be used for this purpose. The 95% confidence interval by the Mid-P exact method is (0.33, 0.96).

Confidence intervals for proportion differences (risk differences and prevalence differences)

This section considers differences between proportions (risk differences and prevalence differences). To avoid redundancy, only the risk difference will be discussed.

Table 13.3 Data for Illustrative Example 13.4. Mortality rates according to physical fitness status, incidence rate data.

Physical fitness improved:

$$\hat{\lambda}_1 = \frac{\text{no} \cdot \text{ disease onsets, exposed}}{\text{sum of person-time, exposed}} = \frac{A_1}{T_1} = \frac{25 \text{ cases}}{4054 \text{ person-years}} = 0.0062 \text{ cases per person-year}$$

Physical fitness not improved:

$$\hat{\lambda}_0 = \frac{\text{no} \cdot \text{ disease onsets, nonexposed}}{\text{sum of person-time, nonexposed}} = \frac{A_0}{T_0} = \frac{32 \text{ cases}}{2937 \text{ person-years}} = 0.0109 \text{ cases per person-year}$$

Source: Blair *et al*. (1995).

Let δ_{risk} denote the risk difference parameter. The **point estimator** of δ_{risk} is

$$\hat{\delta}_{\text{risk}} = \hat{p}_1 - \hat{p}_0 \tag{13.12}$$

where $\hat{p}_1$ represents the incidence proportion in the exposed group and $\hat{p}_0$ represents the incidence proportion in the nonexposed group.

The **standard error of the risk difference** is

$$\text{SE}_{\hat{\delta}_{\text{risk}}} = \sqrt{\frac{A_1 B_1}{N_1^3} + \frac{A_0 B_0}{N_0^3}} \tag{13.13}$$

A **95% confidence interval for δ_{risk}** is calculated with the formula

$$\hat{\delta}_{\text{risk}} \pm (1.96)\,(\text{SE}_{\hat{\delta}_{\text{risk}}}) \tag{13.14}$$

Illustrative Example 13.5 Confidence interval for a risk difference

The data in Table 13.2 have

$$\hat{\delta}_{\text{risk}} = 0.440 - 0.088 = 0.352, \text{SE}_{\hat{d}_{\text{risk}}} = \sqrt{\frac{(11)\,(14)}{25^3} + \frac{(3)\,(31)}{34^3}} = 0.1106$$

and a 95% confidence interval for $\delta_{\text{risk}} = 0.352 \pm (1.96)(0.1106) = (0.14, 0.57)$. An identical result is derived with *OpenEpi.com* → Counts → Two-by-Two Table.

Confidence intervals for rate differences

Let δ_{rate} denote the rate difference parameter. The estimator of this parameter is

$$\hat{\delta}_{\text{rate}} = \hat{\lambda}_1 - \hat{\lambda}_0 \tag{13.15}$$

where $\hat{\lambda}_1$ represents the incidence rate in the exposed group $(= A_1/T_1)$, and $\hat{\lambda}_0$ represents the incidence rate in the nonexposed group $(= A_0/T_0)$.

The **standard error of the rate difference** is

$$\text{SE}_{\hat{\delta}_{\text{rate}}} = \sqrt{\frac{A_1}{T_1^2} + \frac{A_0}{T_0^2}} \tag{13.16}$$

and **the 95% confidence interval for δ_{rate}** is

$$\hat{\delta}_{\text{rate}} \pm (1.96)\left(\text{SE}_{\hat{\delta}_{\text{rate}}}\right) \tag{13.17}$$

Illustrative Example 13.6 Confidence interval for a rate difference

The data in Table 13.3 have a rate difference of $(0.0062 \text{ year}^{-1} - 0.0109 \text{ year}^{-1}) = -0.0047 \text{ year}^{-1}$. The standard error of the rate difference,

$$\text{SE}_{\hat{\delta}_{\text{rate}}} = \sqrt{\frac{25}{4054^2} + \frac{32}{2937^2}} = 0.0023$$

and the 95% confidence interval = −0.0047 ± (1.96)(0.0023) = 0.0047 ± 0.0045 = (−0.0092, −0.0002) years^{-1}. This is equivalent to (−9.2, −0.2) per 1000 person-years. Identical results are derived with *OpenEpi.com* → Person-time → Compare 2 rates.

Confidence intervals for odds ratios, independent samples

The parameter of interest is an incidence odds ratio (cohort study), prevalence odds ratio (cross-sectional study), or exposure odds ratio (case–control study). The notation in Table 13.1 is once again used.

Let ψ represent the odds ratio parameter. The **odds ratio estimator** is simply the cross-product ratio in the 2-by-2 table:

$$\hat{\psi} = \frac{A_1 B_0}{B_1 A_0} \tag{13.18}$$

The **standard error of the *natural logarithm* of the odds ratio** is

$$\text{SE}_{\ln \hat{\psi}} = \sqrt{\frac{1}{A_1} + \frac{1}{B_1} + \frac{1}{A_0} + \frac{1}{B_1}} \tag{13.19}$$

The **95% confidence interval for the ln ψ** is

$$\ln \hat{\psi} \pm (1.96)\left(\text{SE}_{\ln \hat{\psi}}\right) \tag{13.20}$$

Exponents of these limits are the confidence limits for the odds ratio.

Illustrative Example 13.7 Confidence interval for an odds ratio, independent samples

Table 13.4 displays data from a case–control study of esophageal cancer and alcohol consumption. Alcohol consumption is dichotomized (split in two) at 80 g per day. The odds ratio estimate = (96)(666)/(109)(104) = 5.64. The ln($\hat{\psi}$) = ln (5.64) = 1.7299, the standard error of the ln(odds ratio) is

$$\sqrt{\frac{1}{96} + \frac{1}{109} + \frac{1}{104} + \frac{1}{666}} = 0.1752$$

and the 95% confidence interval for ln ψ = 1.7299 ± (1.96)(0.1752) = (1.387, 2.073). The 95% confidence interval for $\psi = e^{(1.387, 2.073)}$ = (4.00, 7.95). An identical result is derived with *OpenEpi.com* → Counts → Two-by-Two Table.

Confidence intervals for odds ratios, matched pairs

Table 13.5 displays the data setup and notation for matched-pair case–control or cohort data. The **odds ratio** estimate here is

$$\hat{\psi} = \frac{u}{v} \tag{13.21}$$

where u represents the number of discordant pairs with an exposed case and v represents the number of discordant pairs with an exposed control. The proportion of discordant pairs in the sample in which the case is exposed is $\hat{p} = u/(u + v)$. Ninety-five

Table 13.4 Data for Illustrative Example 13.7. Esophageal cancer and alcohol consumption, case–control data.

	Case	Control	
≥80 g/day	96	109	205
≤79 g/day	104	666	770
	200	775	975

$$\hat{\psi} = \frac{(96)\,(666)}{(109)\,(104)} = 5.64$$

Source: Tuyns *et al*. (1977) and Breslow and Day (1980).

Table 13.5 Notation for matched-pair data.

	Nondiseased and exposed	Nondiseased and nonexposed
Diseased and exposed	*t*	*u*
Diseased and nonexposed	*v*	*w*

Source: Fleiss (1981, pp. 112–116).

percent confidence limits for proportion p are calculated using Formula (13.3) or an exact method. Call the lower and upper confidence limits for a binomial proportion p $\hat{p}_{\text{LCL}}$ and $\hat{p}_{\text{UCL}}$, respectively. These limits are then transformed into odds ratios as follows:

$$\left(\frac{\hat{p}_{\text{LCL}}}{1-\hat{p}_{\text{LCL}}},\ \frac{\hat{p}_{\text{UCL}}}{1-\hat{p}_{\text{UCL}}}\right) \tag{13.22}$$

(Fleiss, 1981, pp. 112–116). Modern epidemiologic calculators such as WinPepi → WHATIS.EXE → an odds ratio (paired samples) and *OpenEpi.com* → Matched Case Control make this calculation easy.

Illustrative Example 13.8 Confidence interval for an odds ratio, matched pairs

In December of 1980 the CDC published an analysis of risk factors for toxic shock syndrome among tampon users. In this study, cases were matched to ''friend controls.'' The source population was restricted to tampon users. Table 13.6 shows the results for ''continual tampon use'' throughout the menstrual period. The odds ratio estimate = 9/1 = 9.0. The 95% confidence interval for ψ calculated with the help of *OpenEpi.com* → ''Matched Case Control'' is (1.5, 198.9) by the Mid-P exact method.

Table 13.6 Data for Illustrative Example 13.8. Toxic shock syndrome and continual tampon use, matched-pair analysis, matched-pair case–control data.

	Control E +	Control E −
Case E +	33	9
Case E −	1	1

Source: CDC (1980) and Shands *et al*. (1980).

13.3 *p*-Values

> . . . The pitfall is in adopting procedures as things in their own right rather than by having regard to the central objectives the procedures are intended to achieve.
>
> *D.R. Cox, in Armitage (1983, p. 332)*

Hypothesis tests of statistical significance

The basics of interpreting **hypothesis tests** of statistical significance (''significance testing'') was introduced in Section 9.2. Additional comments are provided before several such techniques are introduced.

Two different approaches to statistical testing exist: **significance testing** (Fisher, 1935) and the **fixed-level hypothesis testing** (Neyman and Pearson, 1933). Whether these two techniques represent a unified theory (Lehmann, 1993) or have been improperly combined (Goodman, 1993) is an interesting topic that it is beyond the scope of this text. However, since both methods are used to test statistical hypotheses, and as both lead to substantially the same conclusions, it is convenient to ignore the more subtle distinctions between them for now.

Although confidence intervals are generally preferred in observational studies in epidemiology, significance tests still have a role under some circumstances and still play a central role in experimental studies. However, effective use of this set of techniques requires more than knowing the facts. It requires understanding the underlying reasoning behind the process.

Null and alternative hypotheses

The process of conducting a statistical test begins by specifying a **null hypothesis** ($\boldsymbol{H_0}$). In epidemiology, the null hypothesis is a statement of ''no association between the exposure variable and disease variable.'' As an example, a case–control study might test H_0: $\psi = 1$, where ψ represents the odds ratio parameter. Notice that the hypothesis statement references the parameter and not the statistical estimate. The hypothesis competing with the null hypothesis is the **alternative hypothesis** ($\boldsymbol{H_1}$). An all-purpose two-sided statement of the alternative hypothesis is ''H_0 is false.''

In keeping with the prudent skepticism required in experimental research, the null hypothesis is given the benefit of the doubt until the data convince us otherwise. This is analogous to the presumption of innocence in a criminal trial, in which the defendant is presumed innocent until proven guilty. We then use a suitable probability model to ultimately derive a statistic known as the *p*-value.

Basis of the *p*-value

Assuming the null hypothesis is correct, we specify a probability distribution for a test statistics calculated with data from the sample. For example, we might specify that under H_0, the logarithm (base e) of the risk ratio estimate is normally distributed with an expectation of 0 and a standard deviation of $\text{SE}_{\ln\psi}$ ''SE''. Schematically, this **null distribution** is depicted as a curve like the one drawn in Figure 13.2. The probability of observing a log risk ratio that is as extreme or more extreme than the current ln risk ratio statistic assuming H_0 is true *is* the ***p*-value** for the test. The one-sided

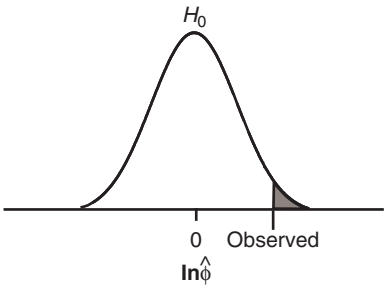


Figure 13.2 Sampling distribution of an risk ratio under the null hypothesis.

p-value corresponds to the area in the tail of the null distribution beyond the observed statistic (shaded region in Figure 13.2). This area is doubled for a two-sided *p*-value. The *p*-value thus quantifies the degree of inconsistency between the data and H_0 that can now be used to address the question ''ought I to take any notice of that?'' (Fisher, 1951, as cited in Lehmann, 1993, p. 1245). A small *p*-value encourages us to say ''yes, the observed association is noteworthy!''[a]

Fallacies of *p*-values and statistical testing

Hypothesis tests of statistical significance are often *misinterpreted*. Here are some common misinterpretations of *p*-values:

- Failure to reject the null hypothesis leads to its acceptance. (Wrong! Failure to reject a false null hypothesis is quite possibly due to insufficient evidence for its rejection.)
- The *p*-value is the probability that the null hypothesis is incorrect. (Wrong! The *p*-value is the probability of the data, assuming the null hypothesis is correct.)
- $p < 0.05$ has an objective basis. (Wrong! $p < 0.05$ is an arbitrary convention that has taken on unwise indiscriminate use.)
- Rejections of H_0 are infallible. (Wrong! False rejections may occur at any level.)
- Small *p*-values provide unassailable support for a causal theory. (Wrong! Without making certain assumptions, *p*-values cannot be used to indicate evidentiary support for a hypothesis.)
- Statistical ''significance'' implies practical importance. (Wrong! Statistical significance is a phrase that has come to mean the null statistical hypothesis has been rejected at a given ''significance level.'')

For observational (nonexperimental) studies, we will view *p*-values as continuous measure of evidence—not merely as significant or insignificant—with smaller and smaller *p*-values providing greater and greater security that the observed association cannot be simply ascribed to chance. Large *p*-values, say, greater than 0.10 or 0.15, indicate that we ought not take too much notice of the observed association.

---

[a] With fixed-level testing, the *p*-value would be compared to a type I error (α) threshold which was set before the experiment was begun (typically 0.05 or 0.01). When the *p*-value falls below this fixed α threshold, the investigator is committed to reject the null hypothesis. The more pliable method described here is more useful when interpreting observational data, encouraging the *p*-value to be viewed in the context of other information.

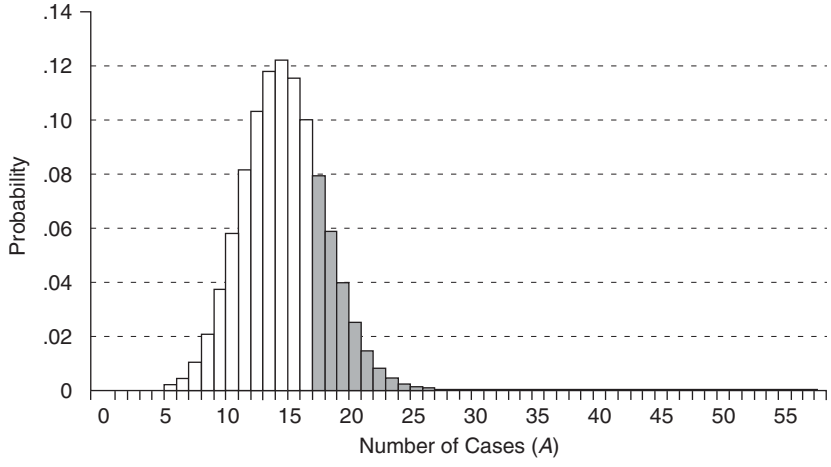


Figure 13.3 Sampling distribution of the number of cases assuming a binomial distribution with $n = 57$ and $p = 0.25$.

Testing a proportion

A binomial test is used to help determine whether a proportion in a given sample is significantly different from what is expected under a null hypothesis. In testing a proportion, let p_0 represent the expected value of the proportion under the null hypothesis. The value of p_0 is borrowed from a previous study or survey, or from the hypothesis itself.

A **normal approximation** to Bernoulli's binomial distribution is used to perform the test when $np_0q_0 \geq 5$, where $q_0 = 1 - p_0$. The test statistic is

$$z = \frac{\hat{p} - p_0}{\text{SE}_{\hat{p}}} \tag{13.23}$$

where $\hat{p}$ is the sample proportion, p_0 is the expected proportion under the null hypothesis, and the **standard error of the proportion** is

$$\text{SE}_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}} \tag{13.24}$$

Illustrative Example 13.9 Hypothesis testing a proportion

A sample of 57 people reveals 17 smokers. Therefore the prevalence of smoking in the sample ($\hat{p}$) = 17/57 = 0.2982, or about 30%. A national survey suggests that we expect about 25% of adults to smoke in this source population ($p_0 = 0.25$). If the source population had a smoking prevalence of 0.25, then the number of smokers in an infinite series of samples will follow a binomial distribution with $n = 57$ and $p = 0.25$. (p in this context refers to the probability parameter of a binomial distribution and not to the *p*-value.) Figure 13.3 displays a binomial distribution with these parameters, which we will refer to as the *null distribution* for the test.

The shaded area in Figure 13.3 represents the one-tailed *p*-value for the test, corresponding to the exact probability of seeing at least 17 cases under the null hypothesis. Because the sample is ''large,''

and np_0q_0 = (57)(0.25)(1 − 0.25) = 10.6875, a normal approximation is used to calculate the area under the curve to derive this *p*-value. (Notice that the distribution shown in Figure 13.3 has a normal ''bell shape''.) For the illustrative data,

$$SE_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.25)\,(1 - 0.25)}{57}} = 0.0574$$

and

$$z = \frac{\hat{p} - p_0}{SE_{\hat{p}}} = z = \frac{0.2982 - 0.25}{0.0574} = 0.84$$

This corresponds to a two-tailed *p*-value of 0.40, suggesting the we ought not pay too much attention to the difference between the observed prevalence in the sample of about 30% and the expected prevalence of 25%. In the jargon that is in common use today, no significant difference was observed.

Testing a rate

The Poisson distribution is used to test whether an observed rate ($\hat{\lambda}$) is significantly different from an expected rate (λ_0). Instead of testing the rate directly, we will direct the test toward the number of cases in the sample. Under the null hypothesis, the number of cases in a given sample is distributed as a Poisson variate with expectation $\mu_0 = \lambda_0 \times T$, where λ_0 represents the null rate derived from a reference population and T represents the sum of person-times in the sample. Calculation of Poisson probabilities is covered in Section 19.2, but is not covered here. Instead, this chapter focuses on setting up the test and calculating the results with WinPEPI.

Illustrative Example 13.10 Hypothesis testing a rate (Poisson random variable)

We observe three cases during 298.5 person-years of observation. Using information from an external reference population, the expected λ_0 is assumed to be 0.003 667 year^{-1}. Therefore, during 298.5 person-years of the observation, the expected number of cases, μ_0 = (0.003 667 year^{-1}) × (298.5 years) = 1.1. Figure 13.4 shows a Poisson distribution assuming μ = 1.1, representing the null distribution for this problem. The *p*-value for this test corresponds to observing *at least* 3 cases when only 1.1 was expected. This corresponds to the shaded area in the tail of the distribution illustrated in Figure 13.4, which sums to 0.0996. Therefore, the one-tailed *p*-value is 0.0996. This *p*-value can be calculated with WinPEPI → Describe → A. Appraise a Rate or Proportion by filling in the observed and expected rates in the clearly labeled fields.

Chi-square test of association

The chi-square test of association is used to test independent *proportions* from cohort studies, cross-sectional studies, and case–control studies. It is one of the most common statistical tests used in epidemiology. The null hypothesis for this test is a statement of ''no association.'' In cohort and cross-sectional studies this may be applied to the risk ratio (H_0: ϕ = 1) or risk difference (H_0: δ = 0). In a case–control study, this is applied to the odds ratio (H_0: ψ = 1).

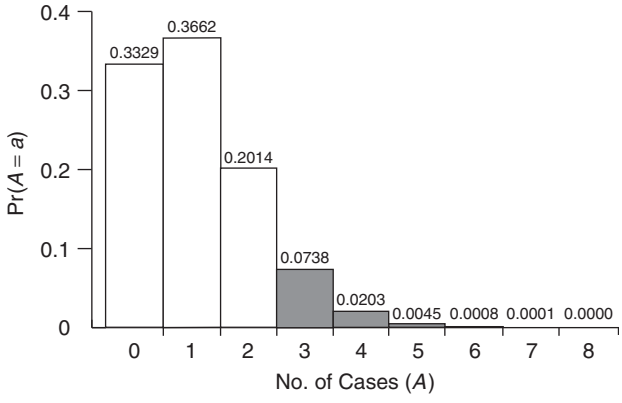


Figure 13.4 Sampling distribution of the random number of cases assuming a Poisson distribution with $\mu = 1.1$.

Table 13.7 Expected frequencies, 2-by-2 cross-tabulations.

	Disease +	Disease −	
Exposure +	$(N_1)(M_1)/N$	$(N_1)(M_0)/N$	N_1
Exposure −	$(N_0)(M_1)/N$	$(N_0)(M_0)/N$	N_0
	M_1	M_0	N

Several different chi-square statistics may be used when conducting this test (e.g., Pearson's uncorrected chi-square statistic, Yates's continuity-corrected chi-square statistic, and the Mantel–Haenszel chi-square statistic). Statisticians do not agree on which of these chi-square statistics is best (Conover, 1974; Mantel, 1974; Miettinen, 1974). Because of its simplicity, we illustrate Pearson's uncorrected chi-square statistic. To calculate this statistic, let O_i represent the observed frequencies in cross-tabulation cell i and let E_i represent the expected frequencies in table cell i. Expected frequencies are calculated as the product of the table's marginal frequencies divided by the total, as shown in Table 13.7. Pearson's **uncorrected chi-square statistic** is

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{13.25}$$

When testing data in a 2-by-2 table, this test statistic has 1 degree of freedom.[b]

[b] The chi-square test statistic can also used to test data in an R-by-C table, in which case the test statistic has $(R - 1)(C - 1)$ degrees of freedom, where R represents the number of rows in the table and C represents the number of columns in the table.

Table 13.8 Data and calculations for Illustrative Example 13.11. Generic cytaribine use and cerebellar toxicity.

Observed frequencies

	Toxicity +	Toxicity −	
Generic +	11	14	25
Generic −	3	31	34
	14	45	59

Expected frequencies

	Disease +	Disease −	
Exposure +	5.93[a]	19.07	25
Exposure −	8.07	25.93	34
	14	45	59

Calculation of chi-square test statistics

$\chi^2 = (11 - 5.93)^2/(5.93) + (14 - 19.07)^2/(19.07) + (3 - 8.07)^2/(8.07)$
$+ (31 - 25.93)^2/(25.93) = 4.33 + 1.35 + 3.19 + 0.99 = 9.85$

df = (2 − 1)(2 − 1) = 1

$p = 0.0017$

[a]Example of calculation: $E_{A1} = (25)(14)/59 = 5.93$.
Source: Jolson *et al*. (1992).

Illustrative Example 13.11 Hypothesis testing independent proportions, chi-square test of association

We submit the data from Illustrative Example 13.3 in Table 13.2 to a chi-square test. The null hypothesis is H_0: $\phi = 1$ (''no association''). The alternative hypothesis H_1 is ''H_0 is untrue.'' Table 13.8 shows the observed and expected frequencies for these data. The chi-square test statistic is calculated below the tables ($\chi^2 = 9.85$, df = 1). The *p*-value of 0.0017 was derived with WinPEPI → WhatIs → P value. The small *p*-value provides strong evidence that the observed association cannot be easily explained by chance.

Note the test statistic is calculated directly with WinPEPI → Compare2 → A. Proportions. It is also provided by *OpenEpi.com* → Counts → Two-by-Two Tables.

Fisher's exact test

Chi-square tests should be avoided when one of the expected frequencies in the 2-by-2 table is less than 5. In such instances, Fisher's exact test should be used. Fisher's exact test is based on summing exact binomial probabilities for permutations of table counts that are equal to or more extreme than the observed results, assuming the null hypothesis is true and the table's margins are fixed. Calculation of Fisher's test statistic is explained in most introductory biostatistics tests and will not be covered here. Fisher's exact test results are routinely provided with WinPEPI → Compare2 → A. Proportions or Odds, and by *OpenEpi.com* → Counts → Two-by-Two Tables.

Table 13.9 Data and calculations for Illustrative Example 13.12. Postoperative sodium polystyrene sulfonate in sorbitol and colonic necrosis.

Observed frequencies

	Necrosis +	Necrosis −	
Exposed	2	115	117
Nonexposed	0	862	862
	2	977	979

Expected frequencies

	Necrosis +	Necrosis −	
Exposed	0.24[a]	116.76	117
Nonexposed	1.76	860.24	862
	2	977	979

[a] Example of calculation: $E_{A1} = (117)(2)/(979) = 0.24$.
Source: Gerstman *et al*. (1992).

Illustrative Example 13.12 Fisher's exact test

A study compared the incidence of colonic necrosis in postoperative patients exposed and not exposed to the drug sodium polystyrene sulfonate in sorbitol. Data are displayed in Table 13.9. Notice that two of the cells have expected frequencies that are less than 5. Therefore, the chi-square test is avoided in favor of Fisher's exact test.

The statistical hypotheses are H_0: $\phi = 1$ (''use of the drug not associated with colonic necrosis'') vs. H_1: ''use of the drug associated with colonic necrosis.'' Fisher's exact test calculated with ''WinPEPI → Compare2 → A. Proportions or Odds'' derive a two-tailed p-value of 0.014. This provides good evidence that the observed association is not easily explained by chance.

Testing independent rates

We want to test the difference in two independent rates. The statistical hypotheses are H_0: $\delta_{\text{rate}} = 0$ vs. H_1: $\delta_{\text{rate}} \neq 0$, where δ_{rate} represents the rate difference parameter $\lambda_1 - \lambda_0$. With moderate to large samples, a z test may be applied. The z statistic is

$$z = \frac{\hat{\delta}_{\text{rate}}}{\text{SE}_{\hat{\delta}_{\text{rate}}}} \tag{13.26}$$

where $\hat{\delta}_{\text{rate}}$ represents the observed rate difference $\hat{\lambda}_1 - \hat{\lambda}_0$ and $\text{SE}_{\hat{\delta}_{\text{rate}}} = \sqrt{\frac{A_1}{T_1^2} + \frac{A_0}{T_0^2}}$ is the standard error of the rate difference.

Illustrative Example 13.13 Hypothesis testing two rates

Recall the fitness and mortality data in Table 13.3. The observed rate difference $\hat{\lambda}_1 - \hat{\lambda}_0 = 0.0062$ year^{-1} $- 0.0109$ year^{-1} $= -0.0047$ year^{-1}, $\text{SE}_{\hat{\delta}_{\text{rate}}} = \sqrt{\frac{A_1}{T_1^2} + \frac{A_0}{T_0^2}} = \sqrt{\frac{25}{4054^2} + \frac{32}{2937^2}} = 0.0023$, and

$z = \dfrac{\hat{\delta}_{\text{rate}}}{\text{SE}_{\hat{\delta}_{\text{rate}}}} = \dfrac{-0.0047}{0.0023} = -2.04$, for a two-sided p-value of 0.041. The small p-value provides good evidence that the negative association is not easily explained by chance.

We can use *OpenEpi.com* → Person-time → Compare 2 rates to derive exact tests results. WinPEPI → Compare2 → D. Rates with person-time denominators can also be used for this purpose. Fisher's exact two-tailed p-value for these data is 0.044.

McNemar's test for matched pairs

McNemar's test is used to test H_0: $\psi = 1$ for matched-pair data. The uncorrected McNemar's chi-square statistic is

$$\chi^2_{\text{McN}} = \frac{(u - v)^2}{u + v} \tag{13.27}$$

See Table 13.5 for notation. Under the null hypothesis, this test statistic has one degree of freedom.

Illustrative Example 13.14 McNemar's test for matched pairs

Recall the matched-pair data on tampon use and toxic shock syndrome presented as Illustrative Example 13.8. Data are in Table 13.6. In this instance, $\chi^2_{\text{McN}} = \dfrac{(9-1)^2}{9+1} = 6.40$, df = 1, $p = 0.011$ (p-value computed from the chi-square statistic with WinPEPI → WhatIs → P-value). The small p-value suggests that the positive association is not easily explained by chance.

The procedure can be performed with WinPEPI → PairsEtc > A. Yes/No dichotomous variable. It can also be conducted with the assistance of *OpenEpi.com* → Counts → Matched Case-Control. For the current example, $p = 0.012$ by the exact mid-P method (two-tailed), which is in close agreement with uncorrected McNemar statistic.

13.4 Minimum Bayes factors

> We are constantly compelled to assess the degree of credence to be accorded to hypotheses on given data.
>
> *Maurice Kendal (1947, p. 178)*

Introduction

The objective of the Bayesian approach that is about to be presented is to determine the extent to which the relation being studied becomes either more or less plausible after considering the current data. This is analogous to the process of updating a physician's prior belief about whether an individual has a disease based on the results of a particular diagnostic procedure. We can show that the predictive value of a positive test result is small when used in ''screening mode,'' but the predictive value of a positive test is high when used in ''confirmatory mode'' (see Section 4.4). Baysian principles may be applied to statistical testing in a similar manner.

Bayes factor

We consider a method of obtaining minimum Bayes factors using standard significance testing statistics (Goodman, 1999, 2001). Using Bayes' theorem we note that the odds of H_0 being true after the test (**posterior odds**) is the function of the odds of H_0 being true before the test (**prior odds**) and a statistic called the Bayes factor:

$$\text{Posterior odds of } H_0 = \text{prior odds of } H_0 \times \text{Bayes factor} \tag{13.28}$$

where

$$\text{Bayes factor} = \frac{\text{probability of data assuming } H_0}{\text{probability of data assuming } H_1}$$

Interpretation of the Bayes factor

The Bayes factor is a ratio of two probabilities that provides a direct measure of evidentiary weight moving us toward or away from the null hypothesis to a stated degree. For example, a Bayes factor of 1 : 10 suggest that ''the results under the null hypothesis are one-tenth as likely as they are under the alternative.'' Or, ''the evidence supports the null hypothesis one-tenth as strongly as it does the alternative.'' Alternatively, ''the odds of the null hypothesis after the study are one-tenth what they were before the study.'' The Bayes factor is well-suited for weighing scientific evidence.

Prior odds

A difficulty with this approach is the subjectivity involved in selecting prior odds of the null hypothesis. The choice of a prior odds of 1 : 1 has the appeal of appearing ''neutral.'' However, we could also argue that a prior odds should be chosen larger than 1 : 1, since scientific research requires skepticism (Berger and Selke, 1987). For our illustrations, we will adopt prior odds of 1 : 1 to represent ''neutrality'' and 9 : 1 to represent ''skepticism.''

Method to calculate a minimum Bayes factor

For tests based on normal error distributions, a minimum Bayes factor can be calculated with the formula:

$$\text{Minimum Bayes factor} = e^{-z^2/2} \tag{13.29}$$

where e is the universal constant and z is the number of standard errors the estimate falls from the null value. This z score can be obtained directly from a z statistic, t statistic, or the square root of the χ^2 statistic. It can also be derived from the p-value by recognizing the relation between p and z (Table 13.10). For example, a two-sided p-value of 0.05 is associated with a z score of 1.96, while a two-sided p-value of 0.01 is associated with a z score of 2.58, and so on.

Illustrative Example 13.15 Minimum Bayes factor

Suppose we calculate $p = 0.05$ (two-sided). This corresponds to $z = 1.96$. Therefore, the minimum Bayes factor = $e^{-1.96^2/2} = 0.15$. An odds of 0.15 can be expressed as 1 to 6.7. For simplicity, let us round this to 1 to 7. Thus, the posterior odds of H_0 is one-seventh its prior odds. Alternatively, we might say that the odds of the null hypothesis being true after the study results was completed with a *p*-value of 0.05 is one-seventh what it was before the study was completed.

To calculate the posterior probability of H_0 being true, we must declare its prior odds. Suppose, for Illustrative Example 13.15, we had decided on a prior odds of H_0 of 1 : 1, corresponding to ''neutrality.'' Then the posterior odds of H_0 would be $1 \times 0.15 = 0.15$. To convert these odds into a probability, use the formula

$$\text{Probability} = \text{odds}/(1 + \text{odds})$$

Thus, a posterior odds of 0.15 corresponds to posterior probability = 0.15/(1 + 0.15) = 0.13. Notice that this does not appear as ''significant'' as the $p = 0.05$ from which it was derived.

Now let us assume additional skepticism in the form of a the prior odds of H_0 of 9 : 1, as might be appropriate in a data dredging exercise. A minimum Bayes factor of 0.15 (derived from $p = 0.05$) in this situation provides posterior odds = $9 \times 0.15 = 1.35$. A posterior odds of 1.35 translates to a posterior probability = 1.35/(1 + 1.35) = 0.57, which is hardly convincing evidence against H_0.

Table 13.10 lists relations between selected *p-v*alues, minimum Bayes factors, and posterior probabilities of H_0 being true when prior odds are 1 : 1 or 9 : 1. This table shows that *p*-values tend to overstate the evidence against the null hypothesis, especially when skepticism is required.

Table 13.10 Relation between *p*-values, minimum Bayes factors, and posterior probabilities assuming prior H_0 odds of 1 : 1 (''neutral'') and 9 : 1 (''moderately skeptical'').

p-Value (*z* score)	Minimum Bayes factor	Posterior probability assuming prior odds for H_0 = 1 : 1	Posterior probability assuming prior odds for H_0 = 9 : 1
0.10 (1.645)	0.26[a]	0.21[b]	0.70[c]
0.05 (1.96)	0.15	0.13	0.57
0.01 (2.58)	0.04	0.035	0.26
0.001 (3.28)	0.005	0.005	0.043

[a]Example of calculation: minimum Bayes factor = $e^{-z^{2/2}} = e^{-1.645^{2/2}} = 0.26$.

[b]Example of calculation: posterior odds = prior odds × Bayes factor = 1 × 0.26 = .26. To convert from odds to probabilities, use the formula $p = o/(1 + o)$. Thus, an odds of 0.26 corresponds to probability = 0.26/(1 + 0.26) = 0.21.

[c]Example of calculation: posterior odds = prior odds × Bayes factor = 9 × 0.26 = 2.34. An odds of 2.34 corresponds to probability = 2.34/(1 + 2.34) = 0.70.

References

Abramson, J. H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations*, **8** (1), 1 *http://www.epi-perspectives.com/content/8/1/1*.

Armitage, P. (1983) Trials and errors: The emergence of clinical statistics. *Journal of Royal Statistical Society. Series A (General)*, **146**, 321–334.

Berger, J.O., and Selke, T. (1987) Testing a point null hypothesis: The irreconcilability of P-values and evidence. *Journal of American Statistical Association*, **82**, 112–122.

Blair, S.N., Kohl, H.W., 3rd,, Barlow, C.E., Paffenbarger, R.S., Jr., Gibbons, L.W., and Macera, C.A. (1995) Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *JAMA*, **273**, 1093–1098.

Breslow, N.E., and Day, N.E. (1980) *Statistical Methods in Cancer Research*. *Volume 1—The Analysis of Case-Control Studies*, International Agency for Research on Cancer, Lyon.

CDC (1980) Toxic shock syndrome—United States. *MMWR*, **29**, 297–299.

Conover, W.J. (1974) Some reasons for not using the Yates continuity correction on 2 × 2 contingency tables. *Journal of American Statistical Association*, **69**, 374–376, 382.

Dean, A.G., Sullivan, K.M., Soe, M.M. (2006) OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2.3.1. *www.OpenEpi.com*, updated 23 June 2011, accessed 6 July 2012.

Fisher, R.A. (1935) The logic of inductive inference. *Journal of Royal Statistical Society*, **98**, 39–54.

Fleiss, J.L. (1981) *Statistical Methods for Rates and Proportions*, 2nd edn, John Wiley & Sons, Inc., New York.

Gerstman, B.B., Kirkman, R., and Platt, R. (1992) Intestinal necrosis associated with postoperative orally administered sodium polystyrene sulfonate in sorbitol. *American Journal of Kidney Diseases*, **20**, 159–161.

Goodman, S.N. (1993) *p* values, hypothesis tests, and likelihood: Implications for epidemiology of a neglected historical debate. *American Journal of Epidemiology*, **137**, 485–496; discussion 497–501.

Goodman, S.N. (1999) Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, **130**, 1005–1013.

Goodman, S.N. (2001) Of *p*-values and Bayes: A modest proposal. *Epidemiology*, **12**, 295–297.

Hurley, S.F., and Kaldor, J.M. (1992) The benefits and risks of mammographic screening for breast cancer. *Epidemiologic Reviews*, **14**, 101–130.

Jolson, H.M., Bosco, L., Bufton, M.G., Gerstman, B.B., Rinsler, S.S., Williams, E., and Peck, C. (1992) Clustering of adverse drug events: Analysis of risk factors for cerebellar toxicity with high-dose cytarabine. *JNCI*, **84**, 500–505.

Lehmann, E.L. (1993) The Fisher, Neyman–Pearson theories of testing hypotheses: One theory or two? *Journal of American Statistical Association*, **88**, 1242–1248.

Mantel, N. (1974) Comment and a suggestion (in reference to Conover, 1974). *Journal of American Statistical Association*, **69**, 378–380.

Miettinen, O. (1974) Comment. *Journal of American Statistical Association*, **69**, 380–382.

Neyman, J., and Pearson, K. (1933) IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of Royal Society of London. Series A.*, **231**, 287–337.

Shands, K.N., Schmid, G.P., Dan, B.B., Blum, D., Guidotti, R.J., Hargrett, N.T., Anderson, R.L.,Hill, D.L., Broome, C.V, Band, J.D., and Fraser, D.W. (1980) Toxic-shock syndrome in menstruating women: Association with tampon use and *Staphylococcus aureus* and clinical features in 52 cases. *New England Journal of Medicine*, **303**, 1436–1442.

Tukey, J.W. (1962) The future of data analysis. *Annals of Mathematical Statistics*, **33**, 1–67.

Tuyns, A.J., Pequignot, G., and Jensen, O.M. (1977) Esophageal cancer in Ille-et-Vilaine in relation to levels of alcohol and tobacco consumption. Risks are multiplying. *Bulletin du Cancer*, **64**, 45–60.

CHAPTER 14
Mantel–Haenszel Methods

14.1 Ways to prevent confounding

14.2 Simpson's paradox

14.3 Mantel–Haenszel methods for risk ratios

- Mixing of effects
- Homogeneity assumption
- Mantel–Haenszel summary risk ratio
  - Notation
- Confidence interval for the Mantel–Haenszel risk ratio
- Mantel–Haenszel test statistic
  - Epidemiologic calculators

14.4 Mantel–Haenszel methods for other measures of association

- Differences between proportions (incidence proportion difference and prevalence difference)
- Odds ratios
- Rate ratios
- Rate differences
- Test statistic for stratified person-time data

Exercise

References

> As a weighted average of the relative risks this formula would, in the illustration given, yield the over-all relative risk.
>
> *Mantel and Haenszel (1959, p. 736)*

14.1 Ways to prevent confounding

The previous chapter used confidence intervals and significance tests to address random error during data analysis. This chapter uses Mantel–Haenszel methods to address the problem of confounding.

Principles of confounding were introduced in Section 9.2. Recall that confounding derives from inherent differences in risk between the exposed and nonexposed groups that would exist even if the study exposure was absent from both groups. Thus, one way to understand confounding is to consider what might have occurred in the exposed group had the exposure been absent, that is, we ask: ''What is the effect of

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

the exposure in the group when it is isolated from all other causes?'' This idea is counterfactual (''counter to fact'') since the group cannot simultaneously be exposed and nonexposed. Nevertheless, this type of thinking is helpful in providing clues about causation when experimentation is not possible.

Before presenting Mantel–Haenszel methods, let us consider various ways to prevent or control for confounding. These include:

- randomization
- restriction
- matching
- stratification
- regression modeling.

Randomization applies only to experimental studies. However, when ethical and feasible to do so, randomization prevents confounding by balancing extraneous factors (i.e., factors other than the exposure being investigated) in the groups being compared. Both measured and unmeasured extraneous factors tend to distribute equally in the exposed and nonexposed groups when the study exposure is randomly assigned. Confounding can, however, enter into a randomized study when, by chance, groups do not balance with respect to extraneous risk factors. For example, a small randomized study may, by chance, have a control group that is on average older or sicker than the treatment group. Any difference in results observed at the end of this small trial would then be confounded by age and severity of illness. However, because of the ''law of large numbers,'' large trials are unlikely to be imbalanced with respect to extraneous risk factors. Increasing the sample size of a randomized trial thus decreases the potential of confounding. In contrast, increasing the size of a nonrandomized study would have *no* effect on confounding.

Restriction is a technique that imposes uniformity through the use of admissibility criteria in the selection of the people studied. By imposing admissibility criteria, exposed and nonexposed groups are made homogenous with respect to restricted variables. When a study base is homogenous with respect to a potentially confounding factor, this factor can no longer confound results. For example, if a study of daily alcohol consumption and lung cancer was homogeneous with respect to smoking—either all nonsmokers *or* all smokers—then smoking could no longer confound results. Thus, restriction is a simple and effective way to prevent confounding in both nonexperimental and experimental studies.

Matching can be an effective means to control for confounding when applied judiciously. In cohort studies, matching refers to the selection of unexposed subjects who are identical (or similar) to the exposed subjects with respect to the distribution of one or more would-be confounding variables. In case–control studies, matching refers to the selection of controls who are identical (or similar) to the case series with respect to one or more variables. Matching on variables can be done on a one-to-one basis (**individual matching**) or can involve matching on factors that define strata (**frequency matching**). Note, however, that individual matching in case–control studies does not control for confounding unless a proper matched analysis is used.

Stratification is a common way to control for confounding in observational studies. This requires the epidemiologist to classify data into homogeneous subgroups and

then use a statistical method to derive an adjusted summary measure of association. This chapter introduces one such method—the Mantel–Haenszel method—just for this purpose.

Regression models are used in epidemiologic studies to evaluate the causal role of one or more exposures while controlling for the confounding effects of other risk factors. Such models are particularly useful when many variables are to be investigated. Care is needed in using regression models, however, because regression models impose assumptions that are not transparent. Thus, even when regression modeling is used, they should be preceded with the type of stratified analysis technique about to be introduced (Vandenbroucke, 1987).

14.2 Simpson's paradox

Simpson's paradox (1951) is a strong form of confounding that results in the reversal of the direction of an association (Rothman, 1975).

Illustrative Example 14.1 Blyth's example of Simpson's paradox

Suppose a doctor is testing a treatment in two separate clinics. A statistician advises the doctor to allocate the treatment so that 91% of the patients in clinic 1 are randomly assigned the new treatment, leaving 9% to the old treatment. In clinic 2, 1% of the patients receive the new treatment, leaving 99% to the old treatment. (Treatments were assigned to provide the appropriate number of patients that could be handled at each location.) Upon completion of the study, the doctor gives the data to the statistician who cross-tabulates the data to form a single 2-by-2 table (Table 14.1A). Based on this analysis, the statistician criticizes the new treatment as a bad one. The doctor is baffled, however, as he perceives the treatment as a good one. The paradox is solved when data are stratified by clinic. Within each clinic, the new treatment is effective, approximately doubling recovery rates at each of the sites (Table 14.1B).

How does one explain these paradoxical effects? ''As with any paradox, there is nothing paradoxical once we see what has happened'' (Blyth, 1972). Patients in clinic 1 were simply much less likely to recover than patients in clinic 2, and the new treatment was given mostly to clinic 1 patients. Therefore, the poor results of the treatment overall merely reflected its propensity for use in the clinic with more severe illness. This bias acted to make the new treatment appear worse than it actually was. If the tendency to use the new treatment was reversed so that patients in clinic 2 were preferentially exposed to it, the bias would have acted in the opposite direction.

14.3 Mantel–Haenszel methods for risk ratios

Mixing of effects

Confounding comes from the mixing of the effects of the confounding variable with the effects of the exposure variable. In the above example, the clinic was a surrogate variable for the seriousness of the illness being treated. Confounding came about because the comparison of the treatment was also a comparison of outcomes in seriously ill patients and less-ill patients. By stratifying the results into the separate clinics, the subgroups that were formed were relatively homogenous with respect to

Table 14.1 Data for Illustrative Examples 14.1–14.3, "Simpson's paradox."

A. Data for both clinics combined ("crude data")

	Alive	Dead	
New treatment	1095	9005	10 100
Standard treatment	5050	5950	11 000
	6145	14 955	

Incidence of recovery, new treatment ($\hat{P}_1$) = 1095/10 100 = 11%

Incidence of recovery, standard treatment ($\hat{P}_0$) = 5050/11 000 = 46%

B. Data for the clinics separately ("stratified data")

Clinic 1

	Alive	Dead	
New treatment	1000	9000	10 000
Standard treatment	50	950	1000
	1050	9950	11 000

$\hat{P}_{1,1}$[a] = 1000/10 000 = 10%

$\hat{P}_{0,1}$ = 50/1000 = 5%

Clinic 2

	Alive	Dead	
New treatment	95	5	100
Standard treatment	5000	5000	10 000
	5095	5000	10 100

$\hat{P}_{1,2}$ = 95/100 = 95%

$\hat{P}_{0,2}$ = 5000/10 000 = 50%

[a] In this notation, the first subscript indicates treatment group (1 = new treatment, 0 = standard treatment) and the second subscript indicates stratum (1 = Clinic 1, 2 = Clinic 2).
Source: Blyth (1972).

the confounding variable. In addition, the treatment was randomized within each clinic and the samples were large, suggesting that the results within clinics were not likely to be confounded.

The analysis of our hypothetical treatment could conceivably end here, with data reported separately for each clinic. However, it is often advantageous to summarize the relation being studied with a single, unconfounded measure of association. This can be accomplished by pooling the unconfounded measures of association within clinics to form a single summary measure of association.

Homogeneity assumption

A single unconfounded measure of association can be calculated by pooling measures of association calculated within homogenous strata. Various methods of pooling exist. In this chapter, we cover a flexible set of such techniques called Mantel–Haenszel methods. To apply Mantel–Haenszel methods judiciously, we must assume that the measures of association within strata are homogeneous. This homogeneity assumption allows us to pool strata-specific measures of association to form a single summary measure that has been adjusted for confounding.

Illustrative Example 14.2 Heterogeneity of the incidence proportion differences

Let us return to the data in Table 14.1B. In clinic 1, the incidence proportion difference $\hat{\delta}_1 = \hat{p}_{1,1} - \hat{p}_{0,1}$ = 10% − 5% = 5%. In clinic 2, the incidence proportion difference $\hat{\delta}_2 = \hat{p}_{1,2} - \hat{p}_{0,2}$ = 95% − 50% = 45%. Thus, incidence proportion differences are *not* homogeneous across strata and pooling of the strata-specific incidence proportion differences should be avoided.

Illustrative Example 14.3 Homogeneity of incidence proportion ratios

Let us now consider the potential to pool the strata-specific incidence proportion ratios for data in Table 14.1B. In clinic 1, the incidence proportion ratio $\hat{\phi}_1 = \frac{\hat{p}_{1,1}}{\hat{p}_{0,1}} = \frac{10\%}{5\%} = 2.0$. In clinic 2, the incidence proportion ratio $\hat{\phi}_1 = \frac{\hat{p}_{1,2}}{\hat{p}_{0,2}} = \frac{95\%}{50\%} = 1.9$. These estimates are ''close enough'' to be described with a single incidence proportion ratio. It is therefore reasonable to pool these strata-specific estimates. One would predict that the summary incidence proportion ratio will fall somewhere between 1.9 and 2.0.

In considering the homogeneity condition, strata-specific measures of association need not be identical in order to be pooled. The pooling procedure allows for some statistical variation in measures of association among strata, and should be thought of as an averaging mechanism of strata-specific measures of association. Like all averages, pooling measures of association will fail to capture the variability of its component parts. However, when is it appropriate to suppress this non-uniformity, the pooled measure of association provides a statistical convenience whose purpose is to draw correct conclusions about the effect of the exposure.

Mantel–Haenszel summary risk ratio

The principle behind the Mantel–Haenszel technique is straightforward. Since the measures of association within homogeneous strata are unconfounded, we combine them in the form of an unconfounded summary measure of association. In Illustrative Example 14.3, it is reasonable to say that the unconfounded incidence proportion ratio is about 2, since patients at both clinics were about twice as likely to recover when given the new treatment compared with the old. The Mantel–Haenszel method merely provides a way to calculate a weighted average of strata-specific risk ratios (Cochran, 1954; Mantel and Haenszel, 1959).

Notation

Measures of association between an exposure and disease for all data combined (without stratification) will be called the **crude measures of association** and will be denoted without a subscript. For example, the crude incidence proportion ratio is represented with the symbol $\hat{\phi}$. For the data in Table 14.1A, $\hat{\phi} = 0.24$. Subscripts will be used to denote measures of association within strata. For example, $\hat{\phi}_1$ will represent the incidence proportion ratio in stratum 1, and $\hat{\phi}_2$ will represent the incidence proportion ratio in stratus 2. For the illustrative data, $\hat{\phi}_1 = 2.0$ and $\hat{\phi}_2 = 1.9$. Additional notational conventions are shown in Table 14.2.

The **Mantel–Haenszel summary incidence proportion (risk) ratio** is the weighted average of strata-specific incidence proportion ratios with weights, $w_k = N_{1k}N_{0k}/N_k$. Thus,

$$\hat{\phi}_{\mathrm{MH}} = \frac{\sum_k w_k A_{1k}/N_{1k}}{\sum_k w_k A_{0k}/N_{0k}} = \frac{\sum_k A_{1k}N_{0k}/N_k}{\sum_k A_{0k}N_{1k}/N_k} \tag{14.1}$$

Table 14.2 Labels and notation for stratified 2-by-2 data, stratum k.[a]

	Diseased (1)	Nondiseased (0)	
Exposed (1)	A_{1k}	B_{1k}	N_{1k}
Nonexposed (0)	A_{0k}	B_{0k}	N_{0k}
	M_{1k}	M_{0k}	N_k

[a] A represents the number of cases and B represents the number of noncases.
The first subscript denotes exposed (1) or nonexposed (0).
The second subscript denotes the strata number $\{k: 1, 2, \ldots, K\}$.
N_k (with a single subscript) represents the number of people in that strata.

(Nurminen, 1981). For the illustrative data,

$$\hat{\phi}_{\text{MH}} = \frac{(1000)\,(1000)\,/\,(11\,000) \;+\; (95)\,(10\,000)\,/\,(10\,100)}{(50)\,(10\,000)\,/\,(11\,000) \;+\; (5000)\,(100)\,/\,(10\,100)} = 1.95$$

As illustrations, in cohort studies in which the incidence proportion is the main measure of occurrence:

The incidence proportion in group 1, strata $k \equiv \hat{p}_{1k} = A_{1k}/\text{N}_{1k}$.

The incidence proportion in group 0, strata $k \equiv \hat{p}_{0k} = A_{0k}/N_{0k}$.

The incidence proportion difference in strata $k \equiv \hat{\delta}_k = \hat{p}_{1k} - \hat{p}_{0k}$.

The incidence proportion ratio in strata $k \equiv \hat{\phi}_k = \dfrac{\hat{p}_{1k}}{\hat{p}_{0k}}$.

Case–control studies follow similar conventions. For example the odds ratio in strata $k \equiv \hat{\psi}_k = \dfrac{A_{1k}B_{0k}}{A_{0k}B_{1k}}$.

Confidence interval for the Mantel–Haenszel risk ratio

The estimated **standard error of the natural log of Mantel–Haenszel incidence proportion ratio** is

$$\text{SE}_{\ln\hat{\phi}_{\text{MH}}} = \sqrt{\frac{\sum\left(M_{1k}N_{1k}N_{0k}/N_k^2 \;-\; A_{1k}A_{0k}/N_k\right)}{\left(\sum\left(A_{1k}N_{0k}\right)/N_k\right)\left(\sum\left(A_{0k}N_{1k}\right)/N_k\right)}} \tag{14.2}$$

(Greenland and Robins, 1985).

For the illustrative example,

$$\text{SE}_{\ln\hat{\phi}_{\text{MH}}} = \sqrt{\frac{\left(1050\cdot 10000\cdot 1000/11000^2 - 1000\cdot 50/11000\right) + \left(5095\cdot 100\cdot 10000/10100^2 - 95\cdot 5000/10100\right)}{\left(1000\cdot 1000/11000 + 95\cdot 10000/10100\right)\times\left(50\cdot 10000/11000 + 5000\cdot 100/10100\right)}} = 0.06963$$

The 95% confidence interval for $\ln\,\phi_{\text{MH}}$ is

$$\ln\hat{\phi}_{\text{MH}} \;\pm\; (1.96)\left(\text{SE}_{\ln\,\hat{\phi}_{\text{MH}}}\right) \tag{14.3}$$

For the illustrative example, the 95% confidence interval for $\ln\,\phi_{MH} = \ln(1.95) \pm (1.96)(0.069\,63) = 0.667\,83 \pm 0.136\,47 = (0.531\,36, 0.804\,30)$.

Exponents of these limits provide limits on the nonlogarithmic scale:

$$e^{(0.531\ 36, 0.804\ 30)} = (1.70, 2.24)$$

Mantel–Haenszel test statistic

A test of association (H_0: $\phi_{MH} = 1$ cohort studies; H_0: $\psi_{MH} = 1$ case–control studies) is carried out with the Mantel–Haenszel test statistic:

$$\chi^2_{MH} = \frac{\left(\sum_k A_{1k} - \sum_k \left(N_{1k}M_{1k}\right)/N_k\right)^2}{\sum_k \left(N_{1k}N_{0k}M_{1k}M_{0k}\right)/\left(N_k^2\left(N_k - 1\right)\right)} \tag{14.4}$$

Under the null hypothesis, this test statistic has a chi-square distribution with 1 degree of freedom.

Applying Formula (14.4) to the illustrative data,

$$\chi^2_{MH}\frac{\left[(1000+95) - (10000\cdot 1050/11000 + 100\cdot 5095/10100)^2\right]}{\begin{array}{l}10000\cdot 1000\cdot 1050\cdot 9950/\left(11000^2\,(11000-1)\right)\\ +100\cdot 10000\cdot 5095\cdot 5005/\left(10000^2\,(10100-1)\right)\end{array}} = 78.46$$

and $p = 8.2 \times 10^{-19}$.

A correction for continuity can be applied to the Mantel–Haenszel test statistic as follows:

$$\chi^2_{MH,\ cont.corrected} = \frac{\left(\left|\sum_k A_{1k} - \sum_k \left(N_{1k}M_{1k}\right)/N_k\right| - 0.5\right)^2}{\sum_k \left(N_{1k}N_{0k}M_{1k}M_{0k}\right)/\left(N_k^2\left(N_k - 1\right)\right)} \tag{14.5}$$

The illustrative data derives $\chi^2_{MH,\ cont.corrected} = 77.\ 59$, $p = 1.3 \times 10^{-18}$.

Epidemiologic calculators

Mantel–Haenszel estimates and related statistics for cohort data based on incidence proportions and case–control studies are calculated with WinPEPI → Compare2 → A. Proportions or Odds. Enter data for stratum 1, click the ''Run'' button, click the ''Next Stratum'' button, and then continue the process for each stratum. Click the ''All strata'' button to compute Mantel–Haenszel and other summary statistics. We can also use OpenEpi → Counts → Two-by-Two Tables → Enter data → ''Add Stratum'' button.

14.4 Mantel–Haenszel methods for other measures of association

Mantel–Haenszel methods may be applied to risk differences, odds ratios, rate ratios, and rate differences. While the method of application remains the same, formulas are adapted to each particular measure of association.

Differences between proportions (incidence proportion difference and prevalence difference)

The **Cochran–Mantel–Haenszel summary risk difference estimate** (Cochran, 1954) is

$$\hat{\delta}_{\text{risk, CMH}} = \frac{\sum_k \left(A_{1k}N_{0k} - A_{0k}N_{1k}\right)/N_k}{\sum_k \left(N_{0k}N_{1k}\right)/N_k} \tag{14.6}$$

This is a weighted average of strata-specific risk differences with weights proportional to $N_{1k}N_{0k}/N_k$.

The **standard error** of this estimate is

$$SE_{\hat{\delta}_{CMH}} = \sqrt{\frac{\sum_k \left(\frac{N_{1k}N_{0k})}{N_k}\right)^2 \left(\frac{A_{1k}B_{0k}}{N_{1k}^2\left(N_{1k}-1\right)} + \frac{A_{0k}B_{1k}}{N_{0k}^2\left(N_{0k}-1\right)}\right)}{\left(\sum_k \frac{N_{1k}N_{0k}}{N_k}\right)^2}} \tag{14.7}$$

(Rothman and Greenland, 1998, p. 271). This formula can be used if every denominator (N_{1k} and N_{0k}) is greater than 1. Sato (1989) provides a formula that can be used when zero denominators are encountered.

The 95% confidence interval for δ_{risk} is

$$\hat{\delta}_{\text{risk}} \pm (1.96)\left(SE_{\hat{\delta}_{\text{risk}}}\right) \tag{14.8}$$

The null hypotheses H_0: $\delta_{\text{risk}} = 0$ can be conducted with the test statistics shown as Formula (14.3 or 14.4). Use and interpretation of Formulas (14.6–14.8) are demonstrated in Section 17.5 in the context of a survival analysis. WinPEPI and *OpenEpi.com* provide these statistics as part of their stratified analysis output for Proportions and Counts (respectively).

Odds ratios

The Mantel–Haenszel summary odds ratio is

$$\hat{\psi}_{\text{MH}} = \frac{\sum_k A_{1k}B_{0k}/N_k}{\sum_k A_{0k}B_{1k}/N_k} \tag{14.9}$$

The standard error of the natural log of this estimate is

$$SE_{\ln\ \hat{\psi}_{\text{MH}}} = \sqrt{\frac{\sum_k G_kP_k}{2\left(\sum_k G_k\right)^2} + \frac{\sum_k \left(G_kQ_k + H_kP_k\right)}{2\left(\sum_k G_k \sum_k H_k\right)} + \frac{\sum_k H_kQ_k}{2\left(\sum_k H_k\right)^2}} \tag{14.10}$$

where
$G_k = A_{1k}B_{0k}/N_k$
$H_k = A_{0k}B_{1k}/N_k$
$P_k = (A_{1k} + B_{0k})/N_k$
$Q_k = (A_{0k} + B_{1k})/N_k$
(Robins *et al.*, 1986).

The 95% confidence interval for the ln ψ_{mh} is computed in the usual fashion:

$$\ln \widehat{\psi}_{MH} \pm (1.96)\left(SE_{\ln \widehat{\psi}_{MH}}\right) \tag{14.11}$$

Exponents of the confidence limits provide the confidence limits for the odds ratio on a nonlogarithmic scale.

Illustrative Example 14.4 Case–control data

Table 14.3 displays fictitious case–control data we will use to illustrate calculations. The crude odds ratio is 4.94. Within age strata the odds ratios are 1.15 and 1.53, respectively, calculations not shown, suggesting possible heterogeneity. Let us assume for now the differences observed in the strata-specific odds ratios can be ascribed to chance and that the common underlying odds ratio can be adequately described with a single summary measure. The Mantel–Haenszel summary odds ratio is 1.25 (95% confidence interval for ψ_{mh}: 0.32, 4.90; calculations shown in Table 14.3). Thus, the strong positive association seen in the crude data no longer exists after controlling for age. The null hypothesis H_0: $\psi_{mh} = 1$ can be tested with Formula (14.3 or 14.4), when needed. Again, these statistics can be calculated with WinPEPI or *OpenEpi.com* as previously described.

Rate ratios

Let ω represent the rate ratio parameter of interest. Table 14.4 lists notation for stratified person-time data. The Mantel–Haenszel summary rate ratio is

$$\widehat{\omega}_{MH} = \frac{\sum_k A_{1k}T_{0k}/T_k}{\sum_k A_{0k}T_{1k}/T_k} \tag{14.12}$$

(Rothman and Boyce, 1979). This is a weighted average of strata-specific rate ratios with weights proportional to $T_{1k}T_{0k}/T_k$.

The standard error of the estimate is

$$SE_{\ln \widehat{\omega}_{NH}} = \sqrt{\frac{\sum\left(M_{1k}T_{1k}T_{0k}/T_k^2\right)}{\left(\sum\left(A_{1k}T_{0k}\right)/T_k\right)\left(\sum\left(A_{0k}T_{1k}\right)/T_k\right)}} \tag{14.13}$$

(Greenland and Robins, 1985).

The 95% confidence interval for the ln ω_{mh} is

$$\ln \widehat{\omega}_{MH} \pm (1.96)\left(SE_{\ln \widehat{\omega}_{MH}}\right) \tag{14.14}$$

Confidence limits for the rate ratio are derived by taking the exponents of these limits.

Table 14.3 Data for Illustrative Example 14.4, case–control data.

A. Crude data

	Case	Control	
Exposed	21	5	26
Not exposed	80	94	174
	101	99	200

Crude odds ratio (ψ) = (21)(94)/(5)(80) = 4.94

B. Data for age groups separately

Young

	Case	Control	
Exposed	20	61	81
Not exposed	2	7	9
	22	68	90

Old

	Case	Control	
Exposed	1	19	20
Not exposed	3	87	90
	4	106	110

$$\hat{\psi}_{MH} = \frac{(20)(7)/90 + (1)(87)/110}{(2)(61)/90 + (3)(19)/110} = 1.25$$

$G_1 = 1.556$ $\quad G_2 = 0.791$
$H_1 = 1.356$ $\quad H_2 = 0.5182$
$P_1 = 0.3$ $\quad P_2 = 0.8$
$Q_1 = 0.7$ $\quad Q_2 = 0.2$

$$\text{Se}_{\ln\hat{\psi}} = \sqrt{\frac{(1.556)(0.3) + (0.791)(0.8)}{2(1.556 + 0.791)^2} + \frac{[(1.556)(0.7) + (1.336)(0.3)] + [(0.791)(0.2) + (0.5812)(0.8)]}{2[(1.556 + 0.791)(1.356 + 0.5182)]} + \frac{(1.356)(0.7) + (0.5182)(0.2)}{2(1.356 + 0.5182)}}$$

95% confidence interval for $\ln\psi_{MH} = \ln(1.25) \pm (1.96)(0.6864) = 0.2231 \pm 1.365 = (-1.142, 1.588)$
95% confidence interval for $\psi_{MH} = e^{(-1.142,\ 1.588)} = (0.32, 4.90)$

Table 14.4 Notation for stratified person-time data, stratum k

	Diseased	Sum of person-time
Exposed (1)	A_{1k}	T_{1k}
Nonexposed (0)	A_{0k}	T_{0k}
Both groups	M_{1k}	T_k

Rate in exposed group, strata k: $\hat{\lambda}_{k1} = \dfrac{A_{1k}}{T_{1k}}$

Rate in nonexposed group, strata k: $\hat{\lambda}_{k0} = \dfrac{A_{0k}}{T_{0k}}$

Rate ratio in strata k: $\hat{\omega} = \dfrac{\hat{\lambda}_{1k}}{\hat{\lambda}_{0k}}$

Rate difference in strata k: $\hat{\delta}_{\text{rate},\ k} = \hat{\lambda}_{1k} - \hat{\lambda}_{0k}$

Illustrative Example 14.5 Incidence rate data

Table 14.5 contains data on breast cancer rates in women who are ''past-users'' of estrogen replacement therapy and for ''never-users.'' Calculations for the Mantel–Haenszel (MH) rate ratio, standard error of the natural logarithm of the MH rate ratio, and the 95% confidence interval for the MH rate ratio are shown in Table 14.5. The Mantel–Haenszel rate ratio estimate $\widehat{\omega}_{MH} = 0.98$ (95% confidence interval for ω: 0.82, 1.17). Similar results can be derived with WinPEPI → Compare2 → D. Rates with person-time denominators, following the data entry for stratified data described previously. *OpenEpi.com* → Compare 2 Rates → Enter Data → Add Stratum also provides a computational solution.

Rate differences

The point estimator for the Mantel–Haenszel adjusted rate difference parameter is

$$\hat{\delta}_{\text{rate, MH}} = \frac{\sum_k \left(A_{1k}T_{0k} - A_{0k}T_{1k}\right)/T_k}{\sum_k \left(T_{1k}T_{0k}\right)/T_k} \tag{14.15}$$

This estimate has the standard error

$$\text{SE}_{\hat{\delta}_{\text{rate}}} = \sqrt{\frac{\sum_k \left(\left(T_{1k}T_{0k}\right)/T_k\right)^2 \left(A_{1k}/T_{1k}^2 + A_{0k}/T_{0k}^2\right)}{\left(\sum_k \left(T_{1k}T_{0k}\right)/T_k\right)^2}} \tag{14.16}$$

The 95% confidence interval for δ_{rate} is

$$\hat{\delta}_{\text{rate}} \pm (1.96)\left(\text{SE}_{\hat{\delta}_{\text{rate}}}\right) \tag{14.17}$$

Mantel–Haenszel rate difference calculations for Illustrative Example 14.5 are shown in Table 14.5. Note that $\hat{\delta}_{MH} = -0.42 \times 10^{-4}$ (95% confidence interval for δ_{rate}: -3.89×10^{-4}, 3.05×10^{-4}).

Test statistic for stratified person-time data

The null hypothesis can be stated as H_0: ''the exposure is independent of disease rates in the population of those who are potentially exposed'' or H_0: $\omega_{MH} = 1$ or H_0: $\delta_{MH} = 0$. The test statistics is

$$\chi^2_{MH} = \frac{\left(\sum_k A_{1k} - \sum_k \left(M_{1k}T_{1k}\right)/T_k\right)^2}{\sum_k \left(M_{1k}T_{1k}T_{0k}\right)/T_k^2} \tag{14.18}$$

(Shore *et al.*, 1976). Calculation of the Mantel–Haenszel chi-square statistic for Illustrative Example 14.5 is shown in Table 14.6 ($\chi^2_{MH} = 0.06$; df = 1, p value = 0.81), confirming no significant differences in the rates after adjusting for age.

Table 14.5 Data and calculations for Illustrative Example 14.5. Prior estrogen replacement therapy and breast cancer incidence, Nurse's Health Study.

A. Crude data (both age groups combined)

	Cases	Person-years
Past users	173	90 762
Never-users	354	186 466
	527	277 228

$\lambda_1 = 173/90\,762 = 0.001\,906$ years^{-1}; $\lambda_2 = 354/186\,466$ years $= 0.001\,898$ years^{-1}

$\hat{\omega} = 0.001\,906$ year$^{-1}/0.001\,898$ year$^{-1} = 1.00$

B. Data for age groups separately

Younger age group[a]

	Cases	Person-years
Past users	173	90 762
Never-users	354	186 466
	527	277 228

Old age group

	Cases	Person-years
Past users	111	51 750
Never-users	194	89 186
	305	140 936

$\lambda_{1,1} = 62/39\,012$ years $= 0.001\,589$ year^{-1}

$\lambda_{0,1} = 160/97\,280$ years $= 0.001\,645$ year^{-1}

$\hat{\omega}_1 = 0.97$

$\lambda_{1,2} = 111/51\,750$ years $= 0.002\,145$ year^{-1}

$\lambda_{2,2} = 194/89\,186$ years $= 0.002\,175$ year^{-1}

$\hat{\omega}_2 = 0.98$

Mantel-Haenszel confidence interval for the rate ratio

$$\hat{\omega}_{MH} = \frac{(62)(97280)/(136292) + (111)(89186)/(140936)}{(160)(39012)/(136292) + (194)(51750)/(140936)} = \frac{114.4954}{117.0326} = 0.98$$

$$1n\hat{\omega}_{MH} = -0.0219$$

$$SE_{\hat{\omega}_{MH}} = \sqrt{\frac{(222)(39012)(97280)/(136292)^2 + (350)(51750) \times (89186)/(140936)^2}{((62)(97280)/136292 + ((111)(89186)/140936)((160) \times (39012))/136292 + ((194)(51.750)/140936)}} = \sqrt{\frac{1162260}{13399.6896}} = 0.09313$$

95% confidence interval for $\ln \omega_{MH} = -0.0219 \pm (1.96)(0.093\,13) = -0.0219 \pm 0.1825 = (-0.2044, 0.1606)$

95% confidence interval for $\ln \omega_{MH} = e^{(-0.2044,\ 0.1606)} = (0.82, 1.17)$

Mantel-Haenszel confidence interval for the rate ratio

$$\hat{\delta}_{MH} = \frac{[(62)(97280) - (160)(39012)]/136292 + [(111)(89186) - (194)(51750)]/140936}{(39012)(97280)/136292 + (51750)(89186)/140936} = -0.42 \times 10^4$$

$$SE_{\hat{\delta}_{MH}} = \sqrt{\frac{((39012)(97280)/136292)^2\,(62/39012^2 + 160/97280^2) + ((51750)(89186)/140936)^2\,(111/51750^2 + 194/89186^2)}{((39012)(97280)/136292 + 51750)(89186)/140936)}} = 1.772 \times 10^{-4}$$

95% confidence interval for $\delta_{MH} = -0.42 \times 10^{-4} \pm (1.96)(1.772 \times 10^{-4}) = -0.42 \times 10^{-4} \pm 3.47 \times 10^{-4} = (-3.89 \times 10^{-4}, 3.50 \times 10^{-4}$

Mantel-Haenszel chi-square statistic

$$\chi^2_{MH} = \frac{((62 + 111) - [(222)(39012)/136292 + (305)(51750)/140936])^2}{(222)(39012)(97280)/136292^2 + (305)(51750)(89186)/140936^2} = 0.06,\ df = 1, p = 0.81$$

[a] Age groups have been merged to simplify hand calculations.

Source: Colditz *et al*. (1990) as reported in Rosner (1995, p. 594).

Table 14.6 Health insurance status in traced and subjects lost-to-follow-up in a birth cohort.

A. Crude data

	Health-care coverage	No health care coverage	
Traced group	46	370	416
Lost-to-follow-up	195	979	1174

B. Data stratified by race

	White			Black		
	Health-care coverage	No health-care coverage		Health-care coverage	No health-care coverage	
Traced group	10	2	12	36	368	404
Lost-to-follow-up	104	22	126	91	957	1048
	114	24	138	127	1325	1452

Exercise

14.1 Table 14.6 contains crude and stratified data for children in a birth cohort who were traced and who were lost-to-follow-up after 5 years (Morrell, 1999).

(A) Based on the crude data, which group had a higher proportion of individuals with health-care coverage, the traced group or the lost-to-follow-up group?

(B) Based on the stratified data, which group had a higher proportion of individuals with health-care coverage?

(C) How can you explain the apparent contradictory answers in part A and part B of this question?

(D) Without calculation, provide a educated guess of the summary (unconfounded) incidence proportion ratio for health-care coverage after controlling for race.

References

Blyth, C. R. (1972) On Simpson's paradox and the sure thing principle. *Journal of American Statistical Association*, **67**, 364–366.

Cochran, W.G. (1954) Some methods for strengthening the common chi-square tests. *Biometrics*, **10**, 417–451.

Greenland, S., and Robins, J.M. (1985) Estimation of a common effect parameter from sparse follow-up data. *Biometrics*, **41**, 55–68.

Mantel, N., and& Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22**, 719–748.

Morrell, C.H. (1999) Simpson's paradox: An example from a longitudinal study in South Africa. *Journal of Statistical Education*, **7**, *www.amstat.org/publications/jse/secure/v7n3*.

Nurminen, M. (1981) Asymptotic efficiency of general noniterative estimators of common relative risk. *Biometrika*, **68**, 525–530.

Robins, J., Breslow, N., and Greenland, S. (1986) Estimators of the Mantel–Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics*, **42**, 311–323.

Rosner, B. (1995) *Fundamentals of Biostatistics*, 4th edn, Duxbury, Belmont, CA.

Rothman, K.J. (1975) A pictorial representation of confounding in epidemiologic studies. *Journal of Chronic Diseases*, **28**, 101–108.

Rothman, K.J., and Boyce, J.D. (1979) *Epidemiologic analysis with a programmable calculator*. U.S. Government Printing Office, Washington, D.C.

Rothman, K.J., and Greenland, S. (1998) *Modern Epidemiology*, 2nd edn, Lippincott-Raven, Philadelphia.

Sato, T. (1989) On the variance estimator for the Mantel–Haenszel risk difference. *Biometrics*, **45**, 1323–1324.

Shore, R.E., Pasternack, B.S., and Curnen, M.G. (1976) Relating influenza epidemics to childhood leukemia in tumor registries without a defined population base: A critique with suggestions for improved methods. *American Journal of Epidemiology*, **103**, 527–535.

Simpson, E.H. (1951) The interpretation of interaction in contingency tables. *Journal of Royal Statistical Society. Series B*, **13**, 238–241.

Vandenbroucke, J.P. (1987) Should we abandon statistical modeling altogether? *American Journal of Epidemiology*, **126**, 10–13.

CHAPTER 15

Statistical Interaction: Effect Measure Modification

15.1 Two types of interaction
- Types of interaction
- Biological interaction
- Statistical interaction

15.2 Chi-square test for statistical

15.3 Strategy for stratified analysis
- Notes

Exercises

References

> All risk indicators, conditional on exposure or non-exposure, modify at least one of the two common epidemiologic measures of effect-risk difference or risk ratio.
>
> *Miettinen (1974, p. 352)*

15.1 Two types of interaction

Types of interaction

Epidemiologists speak of two distinct types of interaction: biological interaction and statistical interaction. Although these concepts share the term ''interaction,'' these are separate phenomena that should *not* be confused. Biological interaction (**biological interdependence**, **causal interaction**, **synergism**) is the interdependent operation of two or more causes to produce or prevent an effect. **Statistical interaction** (effect measure modification, measure of association heterogeneity) refers to a statistical model that does not adequately predict the joint effects of two or more exposures. Whereas a biological interaction describes a property of causality, statistical interaction does *not* have a universal causal interpretation.

Biological interaction

Because most diseases are caused by multiple causal factors acting together, every cause must be viewed in relation to other causal components. Biological interaction

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

is almost *always* present. Take as an example exposure to the polio virus. The effect of exposure to the polio virus depends not only on the virus, but also on the immune status of the individual. In an immune individual, the virus will have no effect. In a susceptible individual, the virus is causal. The interdependence between the effect of the viral agent and the host factor, which we call immunity, is an example of a biological interaction.

As a second example, consider smoking and lung cancer. Smoking causes lung cancer in only a fraction of exposed individuals. This is because genetic factors and perhaps contributing environmental factors are needed to complete a sufficient causal mechanism for lung cancer. Therefore, we say there is a biological interaction between smoking and complementary contributors to lung cancer. (See Section 2.3 for additional discussion about causal interactions).

Knowledge of biological interactions have direct health relevance. When two factors interact biologically, their combined presence has a more potent effect than either of them singly. Consider as an example the biologic interaction between oral contraceptives and smoking. Both cause cardiovascular disease in women. However, their combined use is a much more potent cause than either alone. Oral contraceptive use is therefore discouraged in smokers because of its causal consequences.

Understanding biological interdependences is useful in helping to focus prevention efforts in susceptible populations. For example, the biological dependency between Down syndrome and pregnancies after the age of 40 allows ''older pregnancies'' to be targeted for screening; the dependency between influenza and mortality in older people and in people with pre-existing cardiorespiratory disease allows vaccination programs to be directed toward these groups; the dependency between driving and alcohol consumption motivates programs against drunk driving.

Statistical interaction

Distinct from biological interaction, statistical interaction (**effect measure heterogeneity**, **measure of association heterogeneity**) occurs when a statistical description does not adequately describe the joint effects of two or more exposures. In contrast to biological interaction, statistical interactions are measure of association specific. For instance, a given risk factor may cause statistical interactions with risk differences but not with risk ratios. The data in Table 15.1 will help illuminate this point. This table lists risks of arterial thromboembolic diseases (myocardial infarction, ischemic stroke) in oral contraceptive user and nonusers according to smoking status. The risk ratio in smokers is 2 ($12 \div 6$); the risk ratio in nonsmokers is also 2 ($4 \div 2$). Therefore, the risk ratios are homogenous across smoking categories. When a measure of association is homogeneous according to levels of another factor, we say that statistical interaction is *absent*. When statistical interaction is absent in the risk ratios, risks plotted on logarithmic graph paper will form parallel lines (Figure 15.1).

Let us now use the data in Table 15.1 to consider whether smoking causes interaction in the risk differences. Among smokers, the risk difference (per 10 000) is $12 - 6 = 6$. Among nonsmokers, the risk difference (per 10 000) is $4 - 2 = 2$. Thus, the risk differences are *heterogeneous* across smoking categories and statistical interaction is *present*. When statistical interaction is present in risk differences, risks plotted on arithmetic graph paper will *not* be parallel (Figure 15.2).

Table 15.1 One-year risk (per 10 000) of thromboembolic diseases according to oral contraceptive use and cigarette smoking, hypothetical values.

	Oral contraceptive users	No oral contraceptive use
Smokers	12	6
Nonsmokers	4	2

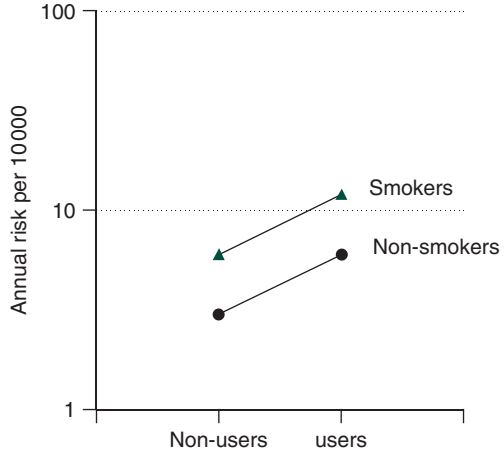


Figure 15.1 One-year risk of thromboembolic disease in oral contraceptive nonusers and users according to smoking status plotted on semi-log paper. Parallel lines indicate homogeneity of the risk ratio.

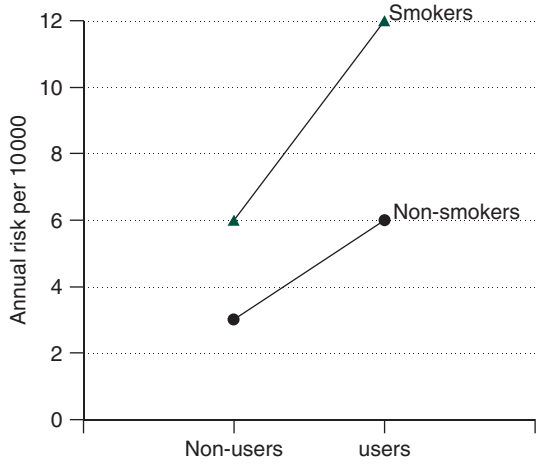


Figure 15.2 One-year risk of thromboembolic disease in oral contraceptive nonusers and users according to smoking status plotted on nonlogarithmic paper. Nonparallel lines indicate heterogeneity of the risk difference.

This illustrates that statistical interaction depends on the measure of association used to describe the relation being studied. Whereas risk ratios in the above example demonstrated no statistical interaction, risk differences demonstrated statistical interaction. Statements about statistical interaction are therefore measure of association specific.

15.2 Chi-square test for statistical

When statistical interaction is present, we expect measures of association to differ in population strata. Since measures of association in epidemiology are also called measures of effect, statistical interaction is also called **effect measure heterogeneity**. Such heterogeneity is often evident upon inspection. However, even in the absence of statistical interaction, effect measures in different strata will vary randomly. To help distinguish random and systematic heterogeneity, the epidemiologist may perform a statistical test for interaction.

Let MA represent the measure of association under consideration. This may be a risk ratio, risk difference, rate ratio, odds ratio, or whatever. We test the following statistical hypotheses:

H_0: $\mathrm{MA}_1 = \mathrm{MA}_2 = \ldots = \mathrm{MA}_K$ (''homogeneity'')

H_1: at least one strata-specific measure of association differs (''heterogeneity'') where MA_k represents measure of association parameter in stratum k { k: 1, 2, ..., K.} For example, in testing odds ratios from two strata, the null and alternative hypotheses are H_0: $\psi_1 = \psi_2$ vs. H_1: $\psi_1 \neq \psi_2$.

An *ad hoc* variance-based chi-square interaction statistic takes the following general form:

$$\chi^2_{\mathrm{INT}} = \sum_{k=1}^{K} \frac{\left(\mathrm{MA}_k - a\mathrm{MA}\right)^2}{\left(\mathrm{SE}_{\mathrm{MA}_k}\right)^2} \tag{15.1}$$

where

$\mathrm{MA}_k \equiv$ measure of association estimate in stratum k

$\mathrm{SE}_{\mathrm{MA}_k} \equiv$ standard error of the measure of effect in stratum k

$a\mathrm{MA} \equiv$ adjusted or summary measures of association (e.g., as a Mantel–Haenszel estimate)

Under the null hypothesis, the χ^2_{INT} statistic has a chi-square distribution with $K-1$ degrees of freedom. Two cautions must be made in reference to this test.

1 Ratio measures of association (odds ratios, risk ratios, rate ratios) are tested on a logarithmic scale. As an example, in testing for heterogeneity of odds ratios, we use $\ln \hat{\psi}_k$ and $\ln \hat{\psi}_{\mathrm{MH}}$ as the MA_k and $a\mathrm{MA}$, respectively. Thus, the interaction statistic for odds ratios is

$$\chi^2_{\mathrm{INT}} = \sum \frac{\ln \hat{\psi}_k - \ln \hat{\psi}_{\mathrm{MH}}}{\left(\mathrm{SE}_{\ln \hat{\psi}}\right)^2_k}$$

2 Difference measures of association (risk differences, rate differences) are tested on a nonlogarithmic scale. However, in testing difference measures of associations, the adjusted measure of association should be based on the maximum-likelihood

estimate (MLE) technique (Rothman and Greenland, 1998, p. 275). Briefly, MLE techniques pool measures of association across strata by determining the most likely value for the parameter given the current data. Calculating MLE statistics requires a computer in most instances. Because of the complexity of MLE calculations we will rely on WinPEPI to compute our interaction statistics.

Illustrative Example 15.1 Interaction in odds ratios (asbestos, lung cancer, and cigarette smoking)

Table 15.2 contains fictitious case–control data describing the relation between asbestos exposure, smoking, and lung cancer. The crude odds ratio (Table 15.2A) is 21.3. Upon stratification (Table 15.2B), we find an odds ratio of 60.0 in smokers and an odds ratio of 2.0 in nonsmokers. This, by itself, provides evidence of statistical interaction. To confirm that the heterogeneity of the odds ratios is nonrandom, we test H_0: $\psi_1 = \psi_2$. Calculations for our *ad hoc* interaction statistic are shown in Table 15.2, deriving a *p*-value of 0.000 022, thus supporting the conclusion of statistical interaction.

WinPEPI → Compare2 → ''stratified input'' derives an odds ratio heterogeneity chi-square statistic of 21.38 with 1 degree of freedom for a *p*-value of 0.000 0038. WinPEPI uses a method attributed to Fleiss (1981, p. 170, formula 10.35) but comes up with much the same conclusion: the large differences in in the strata-specific odds ratios is confirmed as non-random.

Table 15.2 Case–control study of asbestos and lung cancer, fictitious data.

A. Data for smokers and nonsmokers combined

	Case	Control
Exposed	80	38
Not exposed	15	152
	95	190

Crude odds ratio ($\hat{\psi}$) = (80)(152)/(38)(15) = 21.3

B. Data for subgroups

Smokers

	Case	Control
Exposed	75	20
Not exposed	5	80
	80	100

$$\hat{\psi}_1 = \frac{(75)(80)}{(20)(5)} = 60.00$$

$$\ln \hat{\psi}_1 = 4.094$$

$$SE_{\ln} \hat{\psi}_1 = 0.525$$

Nonsmokers

	Case	Control
Exposed	5	18
Not exposed	10	72
	15	90

$$\hat{\psi}_2 = \frac{(5)(72)}{(18)(10)} = 2.00$$

$$\ln \hat{\psi}_2 = 0.693$$

$$SE_{\ln} \hat{\psi}_2 = 0.608$$

Mantel—Haenszel summary odds ratio (calculations not shown)

$\hat{\psi}_{MH} = 16.20$

$\ln \hat{\psi}_{MH} = 2.785$

Ad hoc test for interaction of the odds ratios

H_0: $\psi_1 = \psi_2$ vs. H_1: $\psi_1 \neq \psi_2$

$$\chi^2_{INT} = \sum_K \frac{\ln \hat{\psi}_k - \ln \hat{\psi}_{MH}}{\left(SE_{\ln \hat{\psi}_k}\right)^2} = \frac{(4.094 - 2.785)^2}{0.525^2} + \frac{(0.693 - 2.785)^2}{0.608^2} = 18.05$$

df = $K - 1 = 2 - 1 = 1$

p-value = 0.000 022

15.3 Strategy for stratified analysis

In any epidemiologic study, the investigator thinks carefully about all known factors that may affect the incidence of the outcome being studied. Plans are made in advance for how to measure all such factors and how to address them once data are collected. Whatever the analytic approach, reasoning is made clear. The analysis is flexible yet pragmatic and allows for the detection of unanticipated findings. Mechanical interpretations are avoided.

The analytic strategy about to be presented is intended to prevent the investigator from being overwhelmed by the numerous results that are possible when multiple variables are considered. The strategy is based on uncovering and controlling for hidden confounding and interaction through stratification. Recall from Section 14.2 that findings in the aggregate can show one thing, while findings within strata can reveal something quite different. An expanded illustration will be used to establish the basis of the proposed analytic strategy.

Illustrative Example 15.2 Same crude data, three different outcomes

Table 15.3 contains illustrative data for generic disease D and generic exposure E. The crude risk ratio is 4.0. We are concerned that covariate C (in this case, SEX) might confound the association between E and D, or is involved in a statistical interaction. To simplify this discussion, let us assume data are free of random error, and systematic errors other than might be imparted by factor C are absent as well. Let ϕ_1 represent the risk ratio in stratum 1 (males) and ϕ_2 represent the risk ratio in stratum 2 (females). Upon stratification, data may reveal one of three general patterns:

Scenario A presents a situation that is free of confounding and interaction. In this instance, we note $\phi_1 = 4.0$ and $\phi_2 = 4.0$. Since the strata-specific risk ratios are equal to the crude risk ratio, confounding and statistical interaction are absent. Therefore, there is no practical benefit in stratifying or controlling for the extraneous factor.

Scenario B presents a situation in which confounding is present but statistical interaction is absent. Here, $\phi_1 = 1.0$ and $\phi_2 = 1.0$. Recall that the crude risk ratio was 4.0. Hence, the crude risk ratio was confounded by C. However, since strata-specific risk ratios are uniform, statistical interaction is absent. Under these circumstances, the (unconfounded) risk ratio that describes the relation between E and D can be summarized as a risk ratio of 1.

Scenario C presents a situation in which $\phi_1 = 1.0$ and $\phi_2 = 23.5$. These risk ratios are not uniform. Therefore, statistical interaction is present. Under this circumstance, stratum-specific risk ratios are reported.

Illustrative Example 15.2 presents the following **strategy** for the detection and ''control'' of confounding and interaction (Figure 15.3):

1. Strata-specific measures of association are inspected for evidence of heterogeneity (interaction). A chi-square test for interaction may be used to help confirm that heterogeneity is present. If the investigator concludes statistical interaction is present, measures of association are reported separately for each strata. If statistical interaction is absent or minimal, then proceed to step 2.
2. The investigator calculates a summary measure of association, using a Mantel–Haenszel or alternative technique. If the summary measure of association differs from the crude measure of association in a meaningful way, confounding is likely and the adjusted (summary) measure of association is reported. If evidence of confounding is absent or minimal, then proceed to step 3.

Table 15.3 Data for Illustrative Example 15.2.

Crude data (males and females combined)

	D+	D–	
E+	200	800	1000
E–	50	950	1000

ϕ = (200/1000)/(50/1000) = 4.0

***Scenario A*. No confounding and no statistical interaction**

Males

	D+	D–	
E+	160	240	400
E–	40	360	400

ϕ_1 = (160/400)/(40/400) = 4.0

Females

	D+	D–	
E+	40	560	600
E–	10	590	600

ϕ_2 = (40/600)/(10/600) = 4.0

***Scenario B*. Confounding and no statistical interaction**

Males

	D+	D–	
E+	194	606	800
E–	24	76	100

ϕ_1 = (194/800)/(24/100) = 1.0

Females

	D+	D–	
E+	6	194	200
E–	26	874	900

ϕ_2 = (6/200)/(26/900) = 1.0

***Scenario C*. Statistical interaction**

Males

	D+	D–	
E+	12	188	200
E–	48	752	800

ϕ_1 = (12/200)/(48/800) = 1.0

Females

	D+	D–	
E+	188	612	800
E–	2	198	200

ϕ_2 = (188/800)/(2/200) = 23.5

3 If the crude measure of association and summary measures of association differ only slightly, the likelihood of confounding is small. Under such circumstances, the extraneous variable can be ignored.

Notes

1 Before beginning the study, the investigator learns as much as possible about the complex interrelations among all contributors to the disease being studied. Insights come from previous research, clinical insight, and understanding of the disease process itself. This often requires collaboration with a subject matter specialist.
2 Confounding is a systematic error and is thus *not* amenable to significance testing.
3 Confounders have the following properties: (a) they are associated with the exposure in the source population and (b) they are independent risk factors for the disease. Statistical descriptions of the groups can alert the investigator to the potential for confounding.

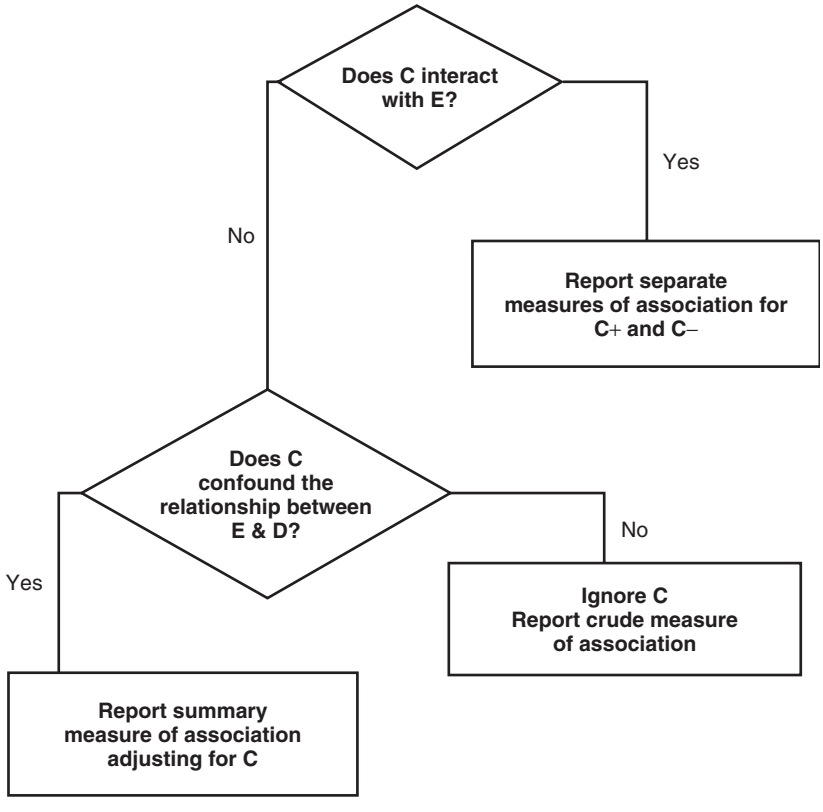


Figure 15.3 A strategy for stratified analysis; E ≡ exposure; D ≡ disease; C ≡ potential confounder.

4 Confounders are not intermediate in the causal pathway between the exposure and disease and are not the consequence of the disease. It is inappropriate to treat intervening causal variables as potential confounders.
5 When interaction and confounding are absent, crude measures of association provide better precision than adjusted and strata-specific measures of association. Thus, the investigator stratifies only on those variables that are necessary to address statistical interactions and control for confounding.
6 In practice, there will always be uncertainty about whether a given set of variables are or are not confounders.

Exercises

15.1 Data from a case–control study of smoking and cervical cancer with data stratified by the number of sexual partners the women had is shown in Table 15.4 (Nischan *et al.*, 1988; Pagano and Gauvreau, 1993).
(A) Calculate the odds ratio for women with zero or one partner.

Table 15.4 Smoking and invasive cervical cancer, case–control study, data for Exercise 15.2.

	Zero to one partner			Two or more partners	
	Case	Control		Case	Control
Smoker	12	21	Smoker	96	142
Nonsmoker	25	118	Nonsmoker	92	150
	37	139		188	292

Source: Nischan *et al*. (1988) as reported by Pagano and Gauvreau (1993, p. 359).

Table 15.5 Current hormone use and breast cancer incidence, Nurses Health Study.[a]

	39–54-year olds			55–64-year olds	
	Cases	Person-years		Cases	Person-years
Current users	85	49 191	Current users	95	26 452
Never-users	160	97 280	Never-users	194	89 186
	245	146 471		289	115 638

[a] Age groups have been merged to simplify calculations.
Source: Colditz *et al*. (1990) as reported by Rosner (1995, p. 594).

(B) Calculate the odds ratio for women with two or one partners.
(C) Based on your analysis so far, do you suspect statistical interaction?
(D) Test H_0: $\psi_1 = \psi_2$.
(E) Would you compute a Mantel–Haenszel summary odds ratio at this point, or would you report odds ratios separately for each stratum? Explain.

15.2 The Nurse's Health Study is a large cohort study in which female nurses were initially mailed questionnaires in 1976 with follow-up questionnaires issued every other year. Data on breast cancer incidence in current users of postmenopausal hormone replacement and never-users are displayed in Table 15.5 (Colditz *et al*., 1990; Rosner, 1995).
(A) Calculate the incidence rate ratio of breast cancer in the 39–54-year old group.
(B) Calculate the incidence ratio of breast cancer in the 55–64-year old group.
(C) Test the rate ratios for heterogeneity (statistical interaction).

References

Colditz, G.A., Stampfer, M.J., Willett, W.C., Hennekens, C.H., Rosner, B., and Speizer, F.E. (1990) Prospective study of estrogen replacement therapy and risk of breast cancer in postmenopausal women. *JAMA*, **264**, 2648–2653.

Fleiss, J.L. (1981) *Statistical Methods for Rates and Proportions*, 2nd edn, John Wiley & Sons, Inc., New York.

Miettinen, O. (1974) Confounding and effect-modification. *American Journal of Epidemiology*, **100**, 350–353.

Nischan, P., Ebeling, K., and Schindler, C. (1988) Smoking and invasive cervical cancer risk. Results from a case–control study. *American Journal of Epidemiology*, **128**, 74–77.

Pagano, M., and Gauvreau, K. (1993) *Principles of Biostatistics*, Duxbury Press, Belmont, CA.

Rosner, B. (1995) *Fundamentals of Biostatistics*, 4th edn, Duxbury Press, Belmont, CA.

Rothman, K. J., nd Greenland, S. (1998) *Modern Eepidemiology*, 2nd edn, Lippincott-Raven, Philadelphia.

CHAPTER 16
Case Definitions and Disease Classification

16.1 Case definitions
- Establishing a case definition
- Multiple-choice criteria
- Chronic fatigue syndrome, as an example
- Evolution of the AIDS case definition, as an example
- Classification of case status based on certainty

16.2 International classification of disease

16.3 Artifactual fluctuations in reported rates

16.4 Summary

References

> Nomenclature is of as much importance [in epidemiology] as weights and measures in the physical sciences.
>
> *William Farr (1885, p. 234)*

16.1 Case definitions

Establishing a case definition

In this chapter, we consider criteria for defining cases and the standardized system of disease nomenclature known as the International Classification of Disease (ICD). We also address how changes in disease nomenclature and coding practices can create artifactual fluctuations in reported rates of disease.

A **case definition** is a set of objective, uniform, and consistent criteria by which to decide whether an individual should be classified as having the disease under investigation. Use of carefully researched and constructed standardized case definitions is important in all epidemiologic endeavors, including epidemiologic surveillance, vital statistics, health surveys, health-care research, descriptive epidemiology, and the various types of experimental and observations analytic studies.

Case definitions may consist of clinical criteria, pathophysiological criteria, and epidemiologic ''person, place, and time'' criteria. For example, a case definition for gastroenteritis associated with a particular food vendor may have criteria that include fever (body temperature greater than or equal to 38.6 °C), diarrhea (bowel

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

movements that conform to the shape of a container), nausea or vomiting, and malaise (self-reported body discomfort and fatigue) in individuals who had eaten at a particular establishment between specific dates. Criteria must be applied consistently throughout the study.

Multiple-choice criteria

Carefully constructed case definition alternatives can be built on multiple criteria. The case definition for coronary heart disease (angina, myocardial infarction, and sudden death) in the Framingham Heart Study (Illustrative Example 7.1), for example, was based on the best clinical and electrocardiographic criteria available at the time as recommended by the New York Heart Association. In cases lacking a clinical history of myocardial infarction, evidence of silent heart attacks were accepted only if an unequivocal pattern of myocardial infarction had developed since the previous electrocardiographic tracing was obtained or there was evidence of prolonged acute coronary insufficiency with electrocardiographic abnormalities (Kannel *et al.*, 1961).

A common approach for building a case definition is to combine diagnostic criteria in different combinations of choices in an ''either/or'' fashion. This is sometimes referred to as a ''Chinese menu''[a] case definition. An example of a Chinese menu case definition for myocardial infarction, used by Henning and Lundman (1975), was to meet two of the criteria listed in A, B, or C (below), *or* criterion D alone.

- Criterion A: central chest pain, pulmonary edema, syncope, or shock.
- Criterion B: pathologic changes detected in the electrocardiogram (e.g., pathological Q-wave).
- Criterion C: two elevated ASAT-values with a maximum approximately 24 h after onset of symptoms in combination with an ALAT-maximum approximately 36 h after onset with ALAT maximum lower than ASAT maximum.
- Criterion D: autopsy findings of myocardial necrosis of an age consistent with the onset of symptoms.

The ''either/or'' aspect of the Chinese menu allows for adjustment to the case definition by either broadening or restricting criteria. A broad case definition is useful during the early phases of an investigation when one needs to gather information on all possible cases. The case definition can later be tightened to allow a sharper focus for testing causal hypotheses, as long as the revised case definition is applied to all past and future cases uniformly.

Chronic fatigue syndrome, as an example

Chronic fatigue syndrome has no known cause and is characterized by chronic fatigue and an array of nonspecific signs and physical symptoms. No laboratory test to confirm the existence of the disease is available, and diagnostic criteria for the syndrome have not been uniformly applied. Researchers, therefore, have lacked adequate tools for assessing the severity and functional limitations of the illness and its response to therapy (Klonoff, 1992). Moreover, without a consistent method for identifying cases,

[a] Historically, Chinese restaurants offered diners combinations of choices, such as ''one from column A or two from column B.''

progress in understanding the epidemiology and clinical correlates of this syndrome have been hampered.

Recognizing these limitations, a group of epidemiologists and academic researchers developed a working case definition based on major and minor clinical and physical criteria (Table 16.1). According to these criteria, a case of chronic fatigue syndrome must fulfill two major diagnostic criteria and eight minor criteria (six or more symptom criteria and two or more of the physical criteria). In 1991, a national workshop updated

Table 16.1 Working case definition of chronic fatigue syndrome.

A case of chronic fatigue syndrome must first fulfill both of the major criteria. In addition, six or more of the symptom criteria plus two or more of the physical criteria must be fulfilled.

Major criteria

1 New onset of persistent or relapsing, debilitating fatigue or easy fatigability in a person with no previous history of similar symptoms. The fatigue must be severe enough to reduce or impair daily activity below 50% of the patient's premorbid activity level for a period of at least 6 months.
2 Other clinical conditions that produce similar symptoms must be excluded thorough evaluation, based on history, physical examination, and appropriate laboratory findings. (Examples of conditions that must be excluded are malignancy, autoimmune disease, localized infection, chronic or subacute bacterial disease, fungal disease, parasitic disease, diseases related to HIV, chronic psychiatric disease, chronic inflammatory disease, neuromuscular disease, endocrine disease, drug dependency or abuse, side effects of chronic medication or other toxic agents, or other known or defined chronic pulmonary, cardiac, gastrointestinal, hepatic, renal, or hematolic diseases.)

Minor criteria

Symptom criteria

1 Chills or mild fever (oral temperature between 37.5 and 38.6 °C, if measured by the patient; oral temperatures greater than 38.6 °C are less compatible with chronic fatigue syndrome and should prompt studies for other causes of illness).
2 Sore throat.
3 Painful lymph nodes in the anterior neck or armpit regions.
4 Unexplained generalized muscle weakness.
5 Muscle pain or myalgia.
6 Prolonged (greater than 24 h in duration) generalized fatigue after modest levels of exercise that would have easily been tolerated in the patient's premorbid state.
7 Generalized headaches (of a type, severity, or pattern different from headaches the patient experienced in the premorbid state).
8 Migratory joint pain without swelling or redness.
9 Neuropsychologic complaints (one or more of the following: photophobia, transient visual scotomata, forgetfulness, excessive irritability, confusion, difficulty thinking, inability to concentrate, or depression)
10 Sleep disturbance (hypersomnia or insomnia).
11 Description of the main symptom complex as initially developed over a few hours to a few days. (This is not a true symptom but may be considered equivalent to the above symptoms in meeting the requirements of the case definition.)

Physical examination criteria

1 Low-grade fever (oral temperature between 37.5 and 38.6 °C or rectal temperature between 37.8 and 38.8 °C).
2 Nonexudative pharyngitis.
3 Palpable or tender anterior or posterior cervical or axillary lymph node (Note: Lymph nodes greater than 2 cm in diameter suggest other causes warranting further evaluation.)

Source: Holmes *et al*. (1988).

this case definition by excluding specific psychiatric diagnoses and postinfectious disease fatigue that could explain the patient's symptoms (Schluederberg *et al*., 1992). Because chronic fatigue syndrome is not a homogeneous abnormality and no single pathogenic mechanism is known, this update further emphasized the need to delineate patient subgroups for separate data analyses. These standards serve as the basis for conducting clinical and epidemiologic studies of chronic fatigue syndrome, providing a rational basis for evaluating patients and discovering the syndrome's cause. (The exact cause of chronic fatigue syndrome is still unknown.)

Evolution of the AIDS case definition, as an example

Case definitions often evolve over time as our understanding of the pathophysiology of disease increases and diagnostic technologies advance. For example, the surveillance case definition of acquired immunodeficiency syndrome (AIDS) has evolved to adapt to our increased understanding of its pathogenesis. Initially, in 1986, the Centers for Disease Control and Prevention (CDC) defined AIDS through the occurrence of a dozen opportunistic infections (e.g., *Pneumocystis carinii* pneumonia) and several cancers (e.g., Kaposi's sarcoma). These diseases—diagnosed by standard clinical, microbiologic, and histopathologic techniques— were considered sufficiently specific to suggest the underlying immunodeficiency, assuming other known causes for the immunodeficiency had been ruled out. In 1987, the CDC revised the surveillance case definition for AIDS to include additional indicator diseases (e.g., wasting syndrome) and to accept as a presumptive diagnosis other indicator conditions if laboratory tests showed concurrent evidence of HIV infection.

The CDC revised the surveillance case definition of AIDS in 1992 to include HIV-infected people who have less than 200 CD4+ T-lymphocytes per microliter of blood or a CD4+ T-lymphocyte percentage of total lymphocytes of less than 14 (Table 16.2). In addition, three new clinical conditions (pulmonary tuberculosis, recurrent pneumonia, and invasive cervical cancer) were added to the surveillance case definition, while retaining the 23 AIDS-defining conditions published by the CDC in 1987. These changes were made to fit the clinical importance of CD4+ lymphocyte level as part of the pathogenesis and medical management of HIV infection. It also simplified identification and reporting of cases and more accurately reflected HIV-related morbidity. A 2008 revision to HIV and AIDS surveillance case definition combined subcategories into a single group and required laboratory-confirmed evidence of HIV infection to meet the case definition among adults, adolescents, and children aged 18 months to <13 years (Schneider *et al*., 2008).

Classification of case status based on certainty

Investigators might choose to classify cases as confirmed, probable, or possible, whenever uncertainty exists. **Confirmed cases** usually require laboratory or special clinical pathologic study verification. **Probable cases** usually have typical clinical features but do not have laboratory or other supporting pathologic confirmation. **Possible cases** have fewer typical clinical features but have a clinical history consistent with the disease in question. CDC *Self-Study Course 3030-G* (CDC, 1992b, p. 358) presents the following examples of confirmed, probable, and possible case definitions for an outbreak of hemolytic-uremic syndrome caused by infection with *Escherichia coli* O157:H7:

- Confirmed case: *E. coli* O157:H7 isolated from a stool culture or development of hemolytic-uremic syndrome in a school-age child resident of the county,

Table 16.2 1993 classification system for HIV infection and expanded AIDS surveillance case definition for adolescents and adults.[a]

CD4+ T-cell count	Clinical categories		
	(A) Asymptomatic, acute (primary) HIV infection or PGL[b]	(B) Symptomatic, not (A) or (C) conditions[c]	(C) AIDS indicator condition[d]
(1) ≥500/μl	A1	B1	C1
(2) 200–499/μl	A2	B2	C2
(3) <200/μl	A3	B3	C3

[a]The shaded cells represent the 1993 AIDS surveillance case definition. People with AIDS indicator conditions (clinical category C) and those with CD4+ T-lymphocyte counts of less than 200/μl are reportable as AIDS cases in the United States. This case definition was revised in 2008 to combine subcategories into a single group and require laboratory confirmation of HIV, demonstrating the evolving nature of case definitions over time.
[b]PGL = persistent generalized lymphadenopathy.
[c]Category B clinical conditions (partial list) include bacillary angiomatosis, mild forms of candidiasis, cervical dysplasia/cervical carcinoma (*in situ*), persistent fever or diarrhea (longer than 1 month in duration), oral hairy leukoplakia, at least two episodes of shingles, idiopathic thrombocytopenia, listeriosis, pelvic inflammatory disease, and peripheral neuropathy.
[d]Category C (AIDS indicator) conditions include severe forms of candidiasis (e.g., of lungs), severe forms of coccidioidomycosis, HIV-related encephalopathy, chronic and severe *herpes simplex* infections, disseminated histoplasmosis Kaposi's sarcoma, several specific forms of lymphoma, several specific forms of *Mycobacterium* infection, pneumocystis pneumonia, recurrent pneumonia, progressive multifocal leukoencephalopathy, recurrent *Salmonella* septicemia, toxoplasmosis of the brain, and HIV-related wasting syndrome.
Source: CDC (1992a, Table 1).

with gastrointestinal symptoms beginning between 3 November and 8 November 1990.

- Probable case: Bloody diarrhea, with the same person, place, and time restrictions as above.
- Possible case: Abdominal cramps and diarrhea (at least three stools in a 24-h period) in a school-age child with onset during the same period as above.

Classifying cases as probable or possible allows the investigator to keep track of potential cases pending confirmation by means of laboratory results. It may also offer economic and practical advantages, especially when dealing with diseases with characteristic clinical pictures (e.g., measles) or when the diagnostic test in question is expensive or difficult to obtain. At times, confirming each case as definite may be unnecessary. For example, in investigating a food-borne outbreak, it is only necessary to isolate the agent from a few afflicted cases. Other descriptive epidemiologic and compatible clinical features will confirm the source and transmission of the agent, making isolation of the agent from each case superfluous.

16.2 International classification of disease

Since 1948, the World Health Organization (WHO) has published the International Classification of Disease (ICD). This scheme provides standardized nomenclature necessary for coding and classifying the causes of morbidity and mortality for regional, national, and international use while helping to enhance diagnostic concordance and reliability. The ICD is currently in its tenth revision (WHO, 1990), although the

ninth revision clinical modification (ICD-9-CM) is still widely used in some clinical environments. The ICD meets the epidemiologist's need for consistency in coding required for rational comparisons of disease trends worldwide.

The ICD9-CM is organized around 17 categories of diseases grouped according to similarities of cause, pathogenesis, and anatomical location. The 17 major categories are:

1 Infectious and Parasitic Diseases (codes 0–139)
2 Neoplasms (140–239)
3 Endocrine, Nutritional, and Metabolic and Immunity Disorders (240–279)
4 Diseases of the Blood and Blood-Forming Organs (280–289)
5 Mental Disorders (290–319)
6 Diseases of the Nervous System and Sense Organs (320–389)
7 Diseases of the Circulatory System (390–459)
8 Diseases of the Respiratory System (460–519)
9 Diseases of the Digestive System (520–579)
10 Diseases of the Genitourinary System (580–629)
11 Complications of Pregnancy, Childbirth, and the Puerperium (630–679)
12 Diseases of the Skin and Subcutaneous Tissue (680–709)
13 Diseases of the Musculoskeletal System and Connective Tissue (710–739)
14 Congenital Anomalies (740–759)
15 Certain Conditions Originating in the Perinatal Period (760–779)
16 Symptoms, Signs, and Ill-Defined Conditions (780–799)
17 Injury and Poisoning (800–999)

Beyond the 17 main categories, supplementary classifications are based on factors influencing health status and contact with health services (V codes) and external causes of injury and poisoning (E codes).

The coding hierarchy of the ICD is achieved by sequencing categories, headings, and subheadings according to clinical detail. For example, the category Diseases of the Circulatory System (390–459) has the following organization:

390–392 Acute rheumatic fever
393–398 Chronic rheumatic heart disease
401–405 Hypertensive disease
410–414 Ischemic heart disease
415–417 Diseases of pulmonary circulation
420–429 Other forms of heart disease
430–438 Cerebrovascular disease
440–448 Diseases of arteries, arterioles, and capillaries
451–459 Diseases of veins and lymphatics, and other diseases of circulatory system

Diseases within subcategories are further organized to provide a basis for indexing and analysis. For example, the individual subcategory Ischemic heart disease (410–414) includes the following three-digit headings:

410 Acute myocardial infarction
411 Other acute and subacute forms of ischemic heart disease
412 Old myocardial infarction
413 Angina pectoris
414 Other forms of chronic ischemic heart disease

Additional specificity is achieved by fourth and fifth digits following a decimal point. For example, the codes under Acute myocardial infarction (410) are:

410.0 Of the anterolateral wall
410.1 Of other anterior wall
410.2 Of inferolateral wall
410.3 Of inferoposterior wall
410.4 Of other inferior wall
410.5 Of other lateral wall
410.6 True posterior wall infarction
410.7 Subendocardial infarction
410.8 Of other specified sites
410.9 Unspecified site

Thus, in this case, ultimate subcategories provide the specific anatomic location of the injury.

16.3 Artifactual fluctuations in reported rates

Artifactual changes in reported morbidity and mortality rates can result from changes in coding and reporting practices. Moreover, the completeness of reported rates varies from study to study and according to region, cause, and other factors. We must be careful when comparing reported rates of disease over time and among studies, especially whenever the ICD is revised and coding practices change.

For example, before 1949, all death certificates that mentioned diabetes as either the immediate cause of death, underlying cause of death, or other significant condition contributing to death were coded as a death due to diabetes. After 1949, this practice changed so that only death certificates listing diabetes as the underlying cause of death were coded as death due to diabetes (Gordis, 1996). This caused an artifactual decline in diabetes death rates (Figure 16.1).

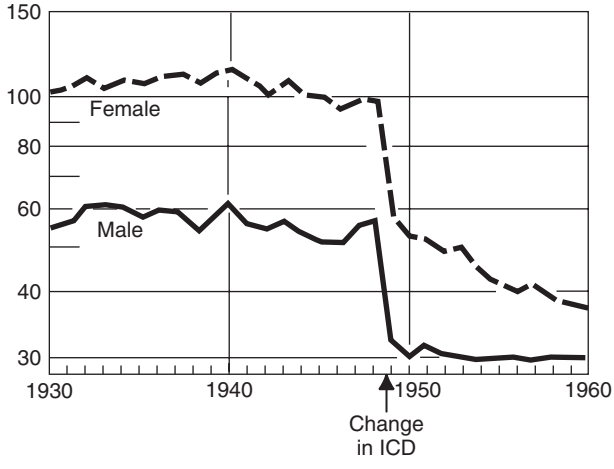


Figure 16.1 Artifactual drops in diabetes death rates in 55–64-year old white men and women, United States, 1930–1960 (Source: National Center for Health Statistics, 1964, p. 36).

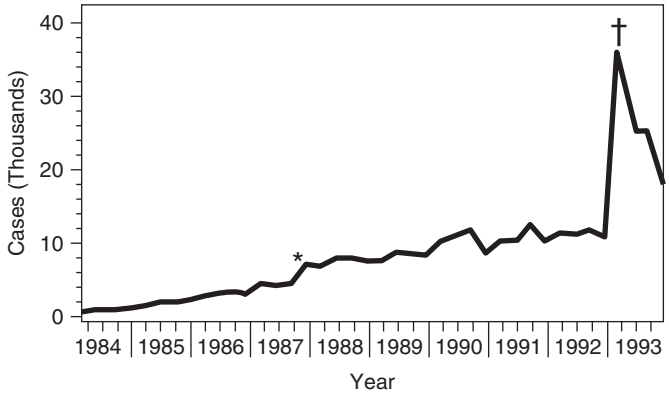


Figure 16.2 AIDS cases by quarter year of report, United States, 1984–1993. *Case definition revised in October 1987 to include additional illnesses and to revise diagnostic criteria. †Case definition revised in 1993 to include CD4+ criteria and three additional illnesses (Source: CDC, 1994, p. 827).

Additional examples of artifactual fluctuations are seen when AIDS surveillance case definitions were changed in 1987 and, again, in 1993. The surveillance case definition was revised in October 1987 to include additional illnesses and diagnostic criteria. In 1993, the surveillance case definition was again changed, this time to include HIV-infected people with CD4+ lymphocyte counts of less than 200 cells/μl but who do not necessarily have an AIDS indicator disease. This resulted in artifactual large influxes in cases with apparent spikes in reporting rates (Figure 16.2).

16.4 Summary

1 The case definition is the standard set of criteria that epidemiologists use for deciding whether an individual should be classified as having the disease or condition under investigation. It is based on uniformly defined objective clinical and epidemiologic criteria. Combinations of criteria can be used to either broaden or restrict the case definition, as dictated by the needs of the investigation. Case definitions evolve over time as our understanding of the pathophysiology of disease increases and diagnostic technologies advance.
2 The International Classification of Disease (ICD) provides a widely accepted, standardized nomenclature for coding and classifying the causes of morbidity and mortality worldwide. It is currently in its tenth revision, although the ninth revision is still used in some environments. The ICD is organized around 17 categories of diseases, grouped according to similarities of cause, pathogenesis, and anatomical location. Consistency in coding is required for rational comparisons of disease trends across various times and regions.
3 Sudden increases or decreases in the reported rates of disease might be due to changes in coding practices or case definitions. Artifactual fluctuations in the rate of a disease are especially likely when the ICD is revised and when surveillance case definitions are altered.

References

Centers for Disease Control and Prevention (CDC) (1986) Classification system for human T-lymphotropic virus type III/lymphadenopathy-associated virus infection. *MMWR*, **35**, 334–339.

Centers for Disease Control and Prevention (CDC) (1987) Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. *MMWR*, **36**, 1S–15S.

Centers for Disease Control and Prevention (CDC) (1992a) 1993 Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR*, **41**, No. RR-17.

Centers for Disease Control and Prevention (CDC) (1992b) *Principles of Epidemiology. Self-Study Course 303-G*, 2nd edn, U.S. Department of Health and Human Services, Atlanta.

Centers for Disease Control and Prevention (CDC) (1994) Update: Trends in AIDS diagnosis and reporting under the expanded surveillance definition for adolescents and adults—United States, 1993. *MMWR*, **43**, 826–827.

Farr, W. (1885) *Vital statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr. (1975 Reprint)*, Offices of the Sanitary Institute, Metuchen, NJ.

Gordis, L. (1996) *Epidemiology*, Philadelphia: W.B. Saunders, Philadelphia.

Henning, R., and Lundman, T. (1975) Swedish co-operative CCU study. A study of 2008 patients with acute myocardial infarction from twelve Swedish hospitals with coronary care units. *Acta Medica Scandinavica*, Supplement 586.

Holmes, G.P., Kaplan, J.E., Gantz, N.M., Komaroff, A.L., Schonberger, L.B., Straus, S.E., *et al*. (1988) Chronic fatigue syndrome: a working case definition. *Annals of Internal Medicine*, **108**, 387–389.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., and Stokes, J., III, (1961) Factors of risk in the development of coronary heart disease—six-year follow-up experience. The Framingham Study. *Annals of Internal Medicine*, **55**, 33–50.

Klonoff, D.C. (1992) Chronic fatigue syndrome. *Clinical Infectious Diseases*, **15**, 812–823.

National Center for Health Statistics (1964) *Vital and Health Statistics. Analytic Studies. The Change in Mortality Trend in the United States*, U.S. Public Health Service Publication No. 1000, Series 3, Number 1. U.S. Government Printing Office, Washington, D.C.

Schluederberg, A., Straus, S.E., Peterson, P., Blumenthal, S., Komaroff, A.L., Spring, S.B., *et al*. (1992) Chronic fatigue syndrome research: Definition and medical outcome assessment. *Annals of Internal Medicine*, **117**, 325–331.

Schneider, E., Whitmore, S., Glynn, K. M., Dominguez, K., Mitsch, A., & McKenna, M. T. (2008). Revised surveillance case definitions for HIV infection among adults, adolescents, and children aged <18 months and for HIV infection and AIDS among children aged 18 months to <13 years–United States, 2008. *MMWR Recomm Rep*, **57**(RR-10), 1–12.

World Health Organization (WHO) (1990) *Tenth revision of the International Classification of Diseases*, World Health Organization, Geneva.

CHAPTER 17
Survival Analysis

17.1 Introduction
17.2 Stratifying rates by follow-up time
17.3 Actuarial method of survival analysis
17.4 Kaplan–Meier method of survival analysis
17.5 Comparing the survival experience of two groups
- Risk differences and risk ratios at selected points in time
- Comparing survival functions as a whole
  - Notes
- Cochran–Mantel–Haenszel chi-square statistic

Exercises
References

> We consider a population of individuals; for each individual we observe either the time to ''failure'' or the time to ''loss'' or censoring.
>
> *D.R. Cox (1972, p. 187)*

17.1 Introduction

Survival analysis encompasses a wide variety of techniques that focus on how long given states ''persist'' over time. This type of analysis has wide application whenever time to onset is important, as is the case in cohort studies and clinical trials.

Survival analysis is particularly important when analyzing data in which risks vary over time. Figure 17.1 displays a survival model for the 1997 US population (solid line). Because mortality increase greatly with age, human survivorship drops off sharply at older ages, creating a survival curve that is ''rectangularized'' in shape. If mortality risk were constant with age, survivorship would demonstrate an exponential decay curve, as demonstrated, for instance, by the dashed line in Figure 17.1. Clearly, constant risk (exponential decay; dashed line) does *not* apply to human survival. In fact, most health risks are not constant over time.

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

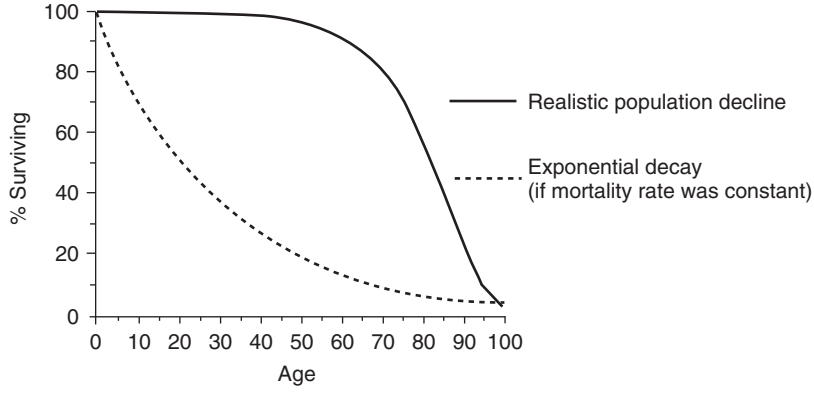


Figure 17.1 Realistic survival curve (solid line) compared with exponential decay (broken line).

Illustrative Example 17.1 Survival data (treatment group)

Let us consider the survival experience of 10 patients treated for a life-threatening disease. Figure 17.2 displays the experience of each patient graphically. The study started in 1990 and ended in 1999. Subjects were enrolled throughout the course of the study. For each study subject, one of the following outcomes was possible:

1 The person died during the period of follow-up.

2 The person withdrew from or discontinued the trial.

3 The person was alive at the completion of the study.

Notice that survival data are complete only for study subjects in category 1. Study subjects in category 2 and category 3 have incomplete survival data in that we do not know the date of their death. Thus, data are truncated or right-censored for these subjects. We combine these right-censored study subjects into a category called **withdrawal**.[a]

The next step in the analysis is to back-up each study subject's follow-up time to ''time zero'' (t_0), the time when they entered the study and, presumably, when they began their treatment (Figure 17.3). The time from a subject's time zero to either withdrawal or death is called the ''person-time of observation'' or simply ''person-time.'' Table 17.1 lists the person-time and outcome for each study subject.

In describing the survival experience of the group, we might initially be tempted to determine the average survival time of the study subjects during the period of study. However, this would substantially underestimate survival by ignoring survival times of study subjects after they withdrew from the study. A more scientific approach calculates the death **rate** in the cohort using the person-time method introduced in Section 3.1. Thus, the observed rate is

$$\hat{\lambda} = \frac{A}{\sum_{N} t_i} = \frac{A}{T} \tag{17.1}$$

[a] An important assumption of standard survival analysis is that withdrawal is independent of survival time.

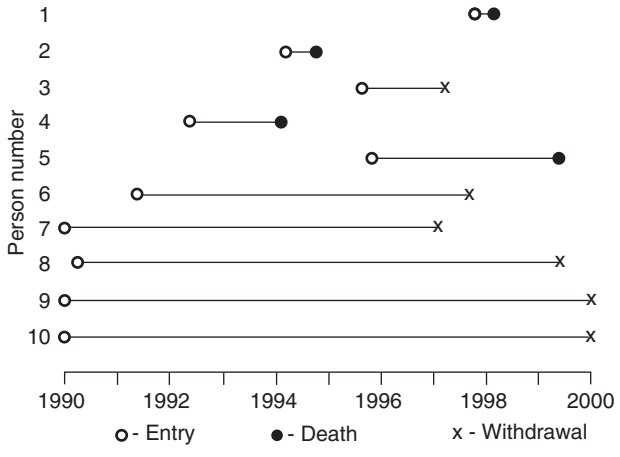


Figure 17.2 Data for Illustrative Example 17.1. Survival experience of treatment group.

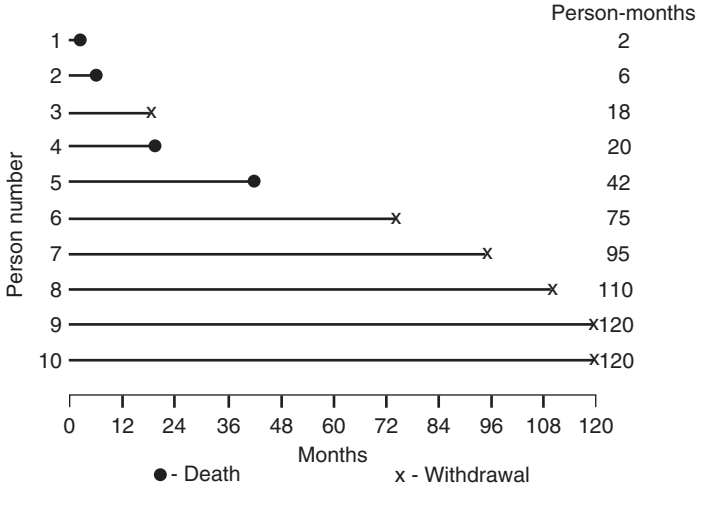


Figure 17.3 Data for Illustrative Example 17.1. Experience of treatment group with data backed up to t_0.

where

$A \equiv$ number of deaths

$t_i \equiv$ person-time for subject i

$T \equiv$ sum of person-time $\left(T = \sum t_i\right)$

For the Illustrative Example 17.1 (Table 17.1), $A = 4$, $T = 608$ months, and $\hat{\lambda} = \dfrac{4}{608 \text{ months}} = 0.006\,58$ per month or, equivalently, 6.58 per 1000 person-months.

The inverse of this rate is the **expected survival time** in the cohort. In this instance, the expected survival time $= \dfrac{1}{\hat{\lambda}} = \dfrac{1}{0.006\,58 \text{ per month}} \cong 152$ months.

Table 17.1 Data for Illustrative Example 17.1. Survival experience of the treatment group.

Subject	Person-months	Outcome
1	2	Death
2	6	Death
3	18	Withdrawal (discontinued study)
4	20	Death
5	42	Death
6	75	Withdrawal (discontinued study)
7	95	Withdrawal (study ended)
8	110	Withdrawal (discontinued study)
9	120	Withdrawal (study ended)
10	120	Withdrawal (study ended)

Although reporting the mortality rate and/or expected survival time is superior to, say, reporting the average survival time during the period of observation, this assumes that the rate of death is constant over time. However, if we carefully examine the illustrative data, we notice that two of the four deaths occurred within 6 months of treatment and all four deaths occurred within 42 months (3.5 years) of treatment. Therefore, hazards are concentrated near the beginning of follow-up. This non-constant hazard needs to be addressed.

17.2 Stratifying rates by follow-up time

One straightforward method for dealing with a non-constant hazard is to stratify rates according to sequential follow-up periods. This is accomplished by grouping the person-time into intervals 1 through K. Rates are then calculated within each interval. Let $\hat{\lambda}_k$ denote the death **rate** in interval k:

$$\hat{\lambda}_k = \frac{A_k}{T_k} \tag{17.2}$$

where A_k is the number of deaths in interval k and T_k is the sum of person-time in that interval.

Let us tally person-time and events within each sequential time-interval. During the first year of follow-up, person 1 contributes 2 person-months of observation time, person 2 contributes 6 person-months, and the remaining 8 people contribute 12 person-months each. Therefore, the sum of person-time during the first year of follow-up, $T_1 = 2 + 6 + (8 \times 12) = 104$ person-months. During this interval, there were 2 deaths (A_1). Consequently,

$$\hat{\lambda}_1 = \frac{2}{104 \text{ months}} = 0.0192 \text{ month}^{-1} = 19.2 \text{ per } 1000 \text{ person-months}$$

During the second year of follow-up, person 1 (now dead) contributes 0 person-months, person 2 (also dead) contributes 0 person-months, person 3 contributes 6

Table 17.2 Data within follow-up intervals, ''modified life table''.

Follow-up interval (k)	Month	No. deaths during interval (A_k)	Sum of person-months during interval (T_k)	Mortality rate per 1000 person-months ($\hat{\lambda}_k$)
1	0–11	2	104	19.2
2	12–23	1	86	11.6
3	24–35	0	72	0.0
4	36–47	1	66	15.2
5	48–59	0	60	0.0
6	60–71	0	60	0.0
7	72–83	0	51	0.0
8	84–95	0	47	0.0
9	96–107	0	36	0.0
10	108–119	0	26	0.0
All intervals combined		4	608	6.6

person-months, person 4 contributes 8 person-months, and the remaining 6 people contribute 6×12 person-months. Thus, the person-time during year 2 of follow-up is $T_2 = 0 + 0 + 6 + 8 + (6 \times 12) = 86$ person-months, during which there was one death. Therefore,

$$\hat{\lambda}_2 = \frac{1}{86 \text{ months}} = 0.0116 \text{ per month} = 11.6 \text{ per } 1000 \text{ person-months}^{b}$$

Table 17.2 lists all of the follow-up interval specific rates. Historically, this type of analysis has been called a **modified life table**.

Notice that in Table 17.2, all of the mortality occurred during the first 4 years of follow-up. The crude mortality rate of 6.6 per 1000 person-months is a weighted average of internal-specific rates. This weighted average fails to capture the non-constant hazard over time. Stratifying the rates into follow-up intervals is a simple way to address this non-constant hazard. It also gives rise to methods for estimating the survival function. Two such methods are the actuarial method and the Kaplan–Meier method.

17.3 Actuarial method of survival analysis

The actuarial method of survival analysis is used to estimate probabilities of death over successive follow-up intervals. Let:

$N_k \equiv$ number of people (''survivors'') entering follow-up interval k

$W_k \equiv$ number withdrawal during interval k

$A_k \equiv$ number of deaths during interval k.

Table 17.3 lists these data elements for the illustrative data in columns (2), (3), and (4), respectively. Notice that the number of people entering interval $k + 1$ is equal to

[b] This is a tedious process when done by hand and is normally done with the computer.

Table 17.3 Actuarial life table, treatment group.

(1) Follow-up interval *(k)*	(2) Entering interval (N_k)	(3) Withdrawals (W_k)	(4) Deaths (A_k)	(5) Effectively exposed to risk (N'_k)	(6) Proportion dying ($\hat{p}_k$)	(7) Proportion surviving ($\hat{q}_k$)	(8) Cumulative proportion surviving to end of interval ($\hat{Q}_k$)
1	10	0	2	10.0	0.2000	0.8000	0.8000
2	8	1	1	7.5	0.1333	0.8667	0.6933
3	6	0	0	6.0	0.0000	1.0000	0.6933
4	6	0	1	6.0	0.1667	0.8333	0.5778
5	5	0	0	5.0	0.0000	1.0000	0.5778
6	5	0	0	5.0	0.0000	1.0000	0.5778
7	5	1	0	4.5	0.0000	1.0000	0.5778
8	4	1	0	3.5	0.0000	1.0000	0.5778
9	3	0	0	3.0	0.0000	1.0000	0.5778
10	3	1	0	2.5	0.0000	1.0000	0.5778

the number entering interval *k* minus the number of withdrawals and deaths in the interval:

$$N_{k+1} = N_k - W_k - A_k \tag{17.3}$$

For the illustrative data, $N_2 = N_1 - W_1 - A_1 = 10 - 0 - 2 = 8$ [column (2), row 2].

The most fundamental information needed to complete an actuarial table is the proportion of people dying within each interval. Before calculating this proportion, we need to compensate for the withdrawals of person-time that occurred during each interval. Therefore, we calculate **the number effectively "exposed" to risk during the interval** (**denoted N'_k**). Several methods may be considered for calculating N'_k. The actuarial method assumes withdrawals occur at mid-interval. This reduces the number of people effectively exposed to risk by half the number of withdrawals:

$$N'_k = N_k - \frac{1}{2}W_k \tag{17.4}$$

For example, in the illustrative data, eight people entered the second interval and one withdrew during this interval. Therefore, $N'_2 = 8 - {}^1/_2(1) = 7.5$. The number of people effectively exposed to risk in the Illustrative Example is listed in column (5) of Table 17.3.

Let $\hat{p}_k$ denote the **incidence proportion of the outcome during interval *k***:

$$\hat{p}_k = \frac{A_k}{N'_k} \tag{17.5}$$

For the illustrative data, $\hat{p}_2 = 1/7.5 = 0.1333$. Values for the other interval-specific incidence proportions are shown in column (6) of Table 17.3.

The **survival proportion in interval *k***, conditional on having survived to that point, denoted $\hat{q}_k$, is the complement of the incidence proportion:

$$\hat{q}_k = 1 - \hat{p}_k \tag{17.6}$$

For the illustrative data, $\hat{q}_2 = 1 - 0.1333 = 0.8667$. Survival proportions are shown in column (7) of Table 17.3.

Comment on notation: We have used p to represent the incidence proportion of the outcome and q to represent its complement (survival proportion) to be consistent with the notation presented earlier in the book. Some sources reverse this convention, using q to represent the proportion dying and p to represent the proportion surviving. Still other sources use R to represent the proportion dying (R stands for risk) and S to represent the proportion surviving.

We are now ready to calculate the **survival function**. Let $\hat{Q}_k$ denote the **cumulative proportion surviving through interval *k***. This is equal to the *product* of the survival proportions up to and including the current interval:

$$\hat{Q}_k = \hat{q}_1\hat{q}_2 \quad \cdots \quad \hat{q}_k = \hat{Q}_{k-1}\hat{q}_k \tag{17.7}$$

For the illustrative data, the cumulative proportion surviving the second year is $\hat{Q}_2 = \hat{Q}_1\hat{q}_2 = (0.8667)\,(0.8000) = 0.6933$. This quantifies the likelihood of surviving the current interval *and* the prior interval. In contrast, the interval-specific survival proportions ($\hat{q}_2$) is *conditional* on having survived all prior intervals. Cumulative survival proportions are listed in column (8) of Table 17.3. This column comprises the **survival function** for the data and is plotted as such in Figure 17.4.

17.4 Kaplan–Meier method of survival analysis

The **Kaplan–Meier product-limit method** (1958) is similar to the actuarial method except for the fact that it places deaths and withdrawals at their precise time, rather than placing them at the middle of the interval in which they occurred. This creates a survival table in which some check points are close together while others are far apart.

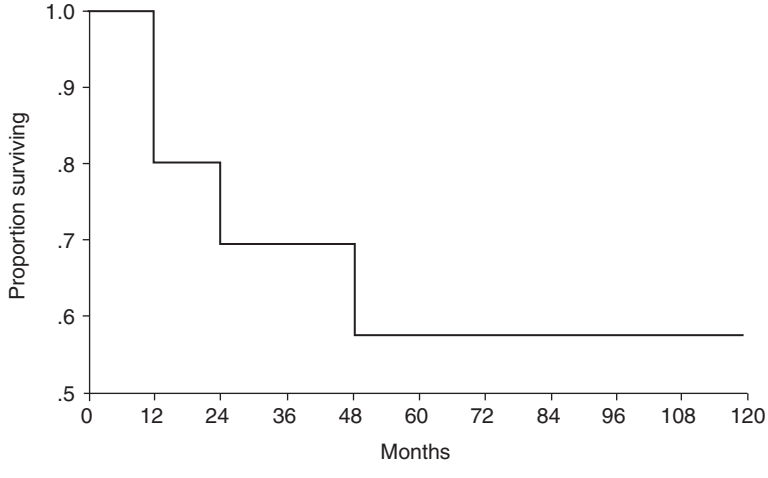


Figure 17.4 Actuarial survival curve, treatment group.

Table 17.4 Kaplan–Meier table, treatment group.

(1) Time of event (month)	(2) Event	(3) Exposed to risk just before event (N_k)	(4) Deaths (A_k)	(5) Proportion dying ($\hat{p}_k$)	(6) Proportion surviving ($\hat{q}_k$)	(7) Cumulative proportion surviving ($\hat{Q}_k$)
2	Death	10	1	1/10	9/10	0.9000
6	Death	9	1	1/9	8/9	0.8000
18	Withdrawal	8	0	0/8	7/7	0.8000
20	Death	7	1	1/7	6/7	0.6857
42	Death	6	1	1/6	5/6	0.5714
75	Withdrawal	5	0	0/5	5/5	0.5714
95	Withdrawal	4	0	0/4	4/4	0.5714
110	Withdrawal	3	0	0/3	3/3	0.5714
120	Withdrawal	2	0	0/2	2/2	0.5714
120	Withdrawal	1	0	0/1	1/1	0.5714

To construct a Kaplan–Meier survival table, follow-up times are listed in rank order, with withdrawals marked with a plus sign (+). The data for our illustrative data (treatment group) are:

2 6 18+ 20 42 75+ 95+ 110+ 120+ 120+

Table 17.4 shows these data in tabular form, with columns for the outcome of each case [column (2)] and the number of subjects effectively exposed to risk just *before* the defining event [column (3)]. The number of deaths at each time point is listed in column (4). Proportion dying ($\hat{p}_k$), proportion surviving ($\hat{q}_k$), and cumulative proportion surviving ($\hat{Q}_k$) are calculated as before, using Formula (17.5), Formula

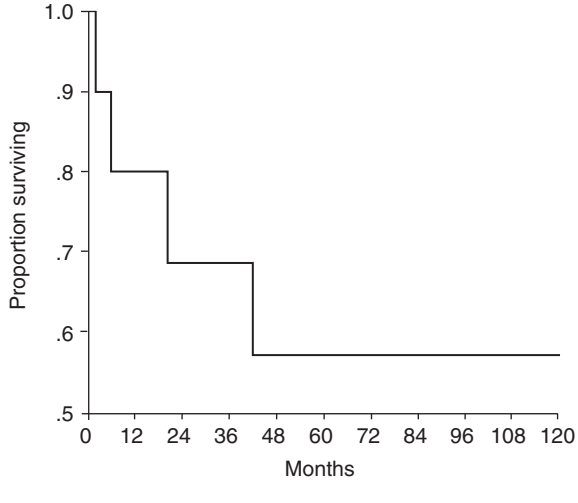


Figure 17.5 Kaplan–Meier survival curve, treatment group.

(17.6), and Formula (17.7), respectively. These statistics appear in columns (5)–(7) of the Table 17.4. Figure 17.5 plots the Kaplan–Meier survival curve for the data.

To derive the Kaplan–Meier survival function (and other survival statistics) using **WinPEPI**[c] (Abramson, 2011), choose WinPEPI → Describe → F. Appraise survival data (time-to-event data).

17.5 Comparing the survival experience of two groups

The survival analysis techniques discussed so far describe the experience of a single group but tell us nothing about how one group's experience sizes up to another. Let us now compare the survival function of the illustrative treatment group with that of a control group.

Illustrative Example 17.2 Survival data (control group)

Let us compare the survival experience of the treatment group we looked at previously with that of a control group. The survival times in months in the control group, with withdrawals marked with a +, are:

4 8 12 18 40+ 60 84 96 108+ 120+

Table 17.5 is a complete Kaplan–Meier survival table for this control group. Figure 17.6 displays the Kaplan–Meier survival curves for the treatment group and control group on the same axis. This graph makes it evident that the survival functions of the two groups overlap until about 60 months, after which the treatment group experiences no further fall-off while the control group continues to dwindle. Although this analysis is based on small numbers, it appears to demonstrate a benefit associated with the treatment.

Table 17.5 Kaplan–Meier life table, control group.

Time of event (month)	Event	At risk before event (N_k)	Deaths (A_k)	Proportion dying ($\hat{p}_k$)	Proportion surviving ($\hat{q}_k$)	Cumulative proportion surviving ($\hat{Q}_k$)
4	Death	10	1	1/10	9/10	0.9000
8	Death	9	1	1/9	8/9	0.8000
12	Death	8	1	1/8	7/8	0.7000
18	Death	7	1	1/7	6/7	0.6000
40	Withdrawal	6	0	0/6	6/6	0.6000
60	Death	5	1	1/5	4/5	0.4800
84	Death	4	1	1/4	3/4	0.3600
96	Death	3	1	1/3	2/3	0.2400
108	Withdrawal	3	0	0/3	2/3	0.2400
120	Withdrawal	3	0	0/3	2/3	0.2400

[c] Download latest version from *www.brixtonhealth.com*.

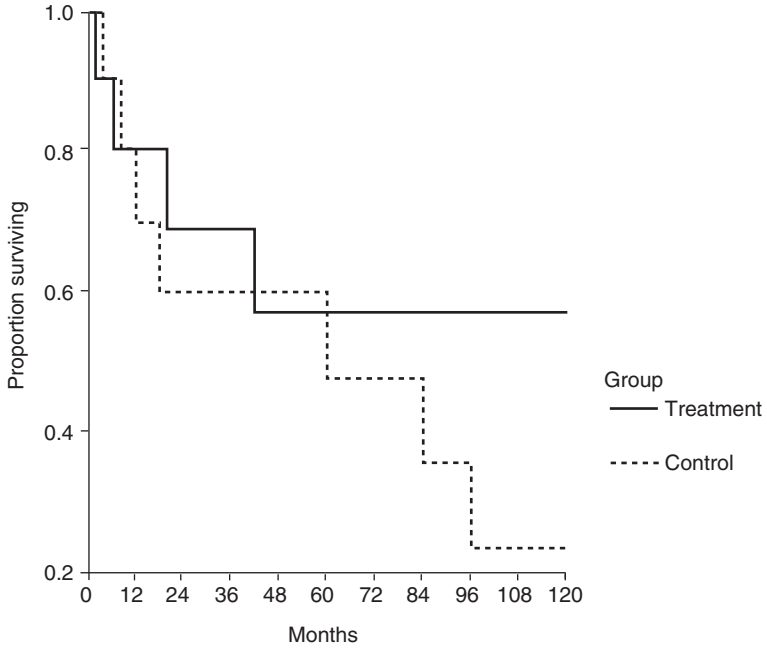


Figure 17.6 Kaplan–Meier survival curve, treatment group (solid line) and control group (broken line).

Risk differences and risk ratios at selected points in time

The relation between treatment and survival can be quantified in terms of a risk difference or risk ratio along selected points of the survival curve. Let:

$\hat{Q}_{1t} \equiv$ cumulative survival proportion in group 1 (treatment) at time t

$\hat{Q}_{0t} \equiv$ cumulative survival proportion in group 0 (control) at time t.

The **cumulative incidence proportion (risk) at time *t*** is the complement of the cumulative survival proportion at that same time mark. In the exposed group (group 1), the cumulative incidence proportion is $\hat{P}_{1t} = 1 - \hat{Q}_{1t}$. In the nonexposed group (group 0), the cumulative incidence proportion (risk) is $\hat{P}_{0t} = 1 - \hat{Q}_{0t}$. Therefore, the **cumulative incidence proportion (risk) difference** at time k is

$$\hat{\delta}_t = \hat{P}_{1t} - \hat{P}_{0t} \tag{17.8}$$

and the **cumulative incidence proportion (risk) ratio** at time t is

$$\hat{\phi}_t = \frac{\hat{P}_{1t}}{\hat{P}_{0t}} \tag{17.9}$$

Illustrative Example 17.3 Risk difference and risk ratio, 5 years in

Five years (60 months) following treatment, the treatment group demonstrates a cumulative survival proportion $\hat{Q}_{1,60\ \text{months}}$ of 0.5714. Therefore, the incidence proportion (risk) at 60 months was $\hat{P}_{1,60\ \text{mo.}} = 1 - \hat{Q}_{1,60\ \text{mo.}} = 1 - 0\ .5714 = 0.4286$. The control group shows a cumulative survival

proportion $\hat{Q}_{0,60\ \text{months}}$ of 0.4800. Therefore, $\hat{P}_{0,60\ \text{mo.}} = 1 - \hat{Q}_{0,60\ \text{mo.}} = 1 - 0.4800 = 0.5200$. The risk difference and risk ratio at 60 months, therefore, are $\hat{\delta}_{60\ \text{mo.}} = \hat{P}_{1,60\ \text{mo.}} - \hat{P}_{0,60\ \text{mo.}} = 0.4286 - 0.5200 = -0.0914$ and $\hat{\phi}_{60\ \text{mo.}} = \dfrac{\hat{P}_{1,60\ \text{mo.}}}{\hat{P}_{0,60\ \text{mo.}}} = \dfrac{0.4286}{0.5200} = 0.82$, respectively.

To derive these and other survival statistics using **WinPEPI**, select Compare2 → H. Numerical observations (including survival times).

Comparing survival functions as a whole

Analysis of cumulative risk ratios and cumulative risk differences requires the investigator to select time points for analysis when many such points are possible. This has the potential to introduce a bias and has the problem of ignoring large sections of data. (The loss of information results in a loss of precision.) An additional problem in choosing points to calculate risk differences involves coming up with a way to address random fluctuations in occurrence over time. To compensate for these problems, we may compare the survival functions of two groups in their entirety. The Mantel–Haenszel procedures introduced in Chapter 14 can be adapted for this purpose (Mantel, 1963; 1966). For example, the **Cochran–Mantel–Haenszel Summary risk difference** (Cochran, 1954) is

$$\hat{\delta}_{\text{CMH}} = \frac{\sum w_k \hat{\delta}_k}{\sum w_k} \tag{17.10}$$

where $\hat{\delta}_k \equiv$ risk difference in interval k $\left(= \hat{p}_1 - \hat{p}_0\right)$ and

$$w_k = \left(N'_{1k}\right)\left(N'_{0k}\right)/N'_k$$

This formula may be rearranged as follows:

$$\hat{\delta}_{\text{CMH}} = \frac{\sum \left(A_{1k}N'_{0k} - A_{0k}N'_{1k}\right)/N_k}{\sum \left(N'_{0k}N'_{1k}\right)/N'_k}$$

(which is the same as Formula 14.6).

Illustrative Example 17.4 Cochran–Mantel–Haenszel summary risk difference (UGDP)

The University Group Diabetes Project (UGDP, 1970) was a long-term study of adult onset diabetic patients that evaluated the effect of various treatments. Table 17.6 summarizes data for the group taking the hypoglycemic agent tolbutamide (group 1) and for the group taking variable amounts of insulin (group 0). Table 17.7 shows calculation of the Cochran–Mantel–Haenszel risk difference statistic. The insulin group averages 1.0% less mortality per 1-year interval.

Table 17.6 Data for Illustrative Example 17.4. UGDP data, tolbutamide versus variable amounts of insulin.

Tolbutamide							
Year (k)	N_{1k}	W_{1k}	A_{1k}	N'_{1k}	$\hat{p}_{1k}$	$\hat{q}_{1k}$	$\hat{Q}_{1k}$
0–(1)	204	0	0	204	0.0000	1.0000	1.0000
1–(2)	204	0	5	204	0.0245	0.9755	0.9755
2–(3)	199	0	5	199	0.0251	0.9749	0.9510
3–(4)	194	5	5	191.5	0.0261	0.9739	0.9262
4–(5)	184	24	5	172	0.0291	0.9709	0.8992
5–(6)	155	41	4	134.5	0.0297	0.9703	0.8725
6–(7)	110	47	5	86.5	0.0578	0.9422	0.8221
7–(8)	58	33	1	41.5	0.0241	0.9759	0.8022
Insulin, variable amounts							
Year (k)	N_{ok}	W_{ok}	A_{ok}	N'_{0k}	$\hat{p}_{0k}$	$\hat{q}_{0k}$	$\hat{Q}_{0k}$
0–(1)	204	0	4	204	0.0196	0.9804	0.9804
1–(2)	200	0	3	200	0.0150	0.9756	0.9565
2–(3)	197	0	3	197	0.0152	0.9848	0.9419
3–(4)	194	5	1	191.5	0.0052	0.9948	0.9370
4–(5)	188	21	0	177.5	0.0000	1.0000	0.9370
5–(6)	167	38	4	148	0.0240	0.9730	0.9117
6–(7)	125	60	1	95	0.0080	0.9895	0.9021
7–(8)	64	35	2	46.5	0.0313	0.9570	0.8633

Data from Elandt-Johnson and Johnson (1980, p. 250).

Table 17.7 Calculation of Illustrative Example 17.4. Cochran–Mantel–Haenszel summary risk difference (UGDP).

k	w_k [a]	$\hat{\delta}_k$ [b]	$w_k\hat{\delta}_k$
1	102.00	−0.0196	−2.0000
2	100.99	0.0095	0.9604
3	99.00	0.0099	0.9798
4	95.75	0.0209	2.0000
5	87.35	0.0291	2.5393
6	70.46	0.0027	0.1912
7	45.28	0.0473	2.1405
8	21.93	−0.0189	−0.4148
SUMS	622.76		6.3964

[a] Weighting factors: $w_k = (N'_{1k})(N'_{0k}) / (N'_k)$; for example, $w_1 = (204)(204)/(204 + 204) = 102.0$.
[b] Stratum specific risk differences $\hat{\delta}_k = \hat{p}_1 - \hat{p}_2$; e.g., $\hat{\delta}_1 = 0.0000 - 0.0196 = -0.0196$.

$$\hat{\delta}_{CMH} = \frac{6.3964}{622.76} = 0.010.$$

Table 17.8 Notation for cross-tabulated data, follow-up interval k[a].

	Death +	Death −	
Group 1	A_{1k}	B_{1k}	N_{1k}
Group 0	A_{ok}	B_{ok}	N_{ok}
	M_{1k}	M_{ok}	N_k

[a]The first subscript denotes group membership; the second subscript denotes follow-up interval.
Risk in the exposed group, interval k: $\hat{p}_{1k} = A_{1k}/N_{1k}$.
Risk in nonexposed group, interval k: $\hat{p}_{0k} = A_{0k}/N_{0k}$.
Risk difference, interval k: $\hat{\delta}_k = \hat{p}_{1k} - \hat{p}_{0k}$
Risk ratio, interval k: $\hat{\phi}_k = \hat{p}_{1k}/\hat{p}_{0k}$.

Notes

1 The standard error of the estimate and confidence interval for δ_{CMH} are calculated with Formulas (14.7 and 14.8), respectively, as needed.
2 When withdrawal is independent of group membership, the adjustment for withdrawals (converting N_{gk} into N'_{gk}) is unnecessary (Mantel, 1966, pp. 164, 168).
3 The survival data from the two groups can be rearranged to form multiple (K) 2-by-2 tables. Notation is shown in Table 17.8.
4 Statistics for Mantel–Haenszel summary risk ratio statistics can be calculated with Formulas (14.1, 14.2, and 14.3).

Cochran–Mantel–Haenszel chi-square statistic

A chi-square statistic is used to test whether there is a significant difference in the survival functions. The **Cochran–Mantel–Haenszel chi-square test statistic** is

$$\chi^2_{\text{CMH}} = \frac{\left(\sum A_{1k} - \sum E_{1k}\right)^2}{\sum V_{1k}} \tag{17.11}$$

where

A_{1k} = observed number of cases in group 1 during interval k
$E_{1k} \equiv$ expected number of exposed cases in group 1 during interval k, calculated

$$E_{1k} = \frac{N'_{1k} M_k}{N'_k}$$

$V_{1k} \equiv$ variance in the expected number of cases in group 1, interval k calculated as

$$V_{1k} = \frac{N'_{1k} N'_{0k} M_{1k} M_{0k}}{N'^2_k \left(N'_k - 1\right)}$$

where $M_{1k} = N'_{1k} - A_{1k}$ and $M_{0k} = N'_{0k} - A_{1k}$ This test statistic has one degree of freedom.

Table 17.9 Calculation of Cochran–Mantel–Haenszel chi-square statistic, UGDP data, tolbutamide vs. variable insulin.

k	A_{1k}[a]	E_{1k}[b]	V_{1k}[c]
1	0	2.00	0.99
2	5	4.04	1.97
3	5	4.02	1.96
4	5	3.00	1.48
5	5	2.46	1.24
6	4	3.81	1.95
7	5	2.86	1.46
8	1	1.41	0.73
SUMS	30	23.60	11.77

[a]Observed number of cases in group 1 during interval k.
[b]Expected number of cases in group 1 during interval k; for example,

$$E_{1,1} = \frac{(204)\,(4)}{(204 + 204)} = 2.00.$$

[c]Variance in the expected number of cases in group 1, interval k; for example,

$$V_{1,1} = \frac{(204)\,(204)\,(4)\,(404)}{408^2\,(408 - 1)} = 0.99.$$

Illustrative Example 17.5 Cochran–Mantel–Haenszel chi-square statistic (UGDP)

Table 17.9 shows the components of the chi-square calculation for the UGDP data. The resulting p-value of 0.062 suggests that the observed difference in survival in the UGDP trial (Illustrative Example 17.4) is not easily explained by chance.

$$\chi^2_{CMH} \frac{(30 - 23.60)^2}{11.77} = 3.48;\ \text{df} = 1;\ p = .062.^{d}$$

Exercises

17.1 Data for this exercises come from a study by Crowley and Hu (1974) reported in Elandt-Johnson and Johnson (1980, pp. 159–160). Data may be downloaded from the *Epidemiology Kept Simple* website data directory in SPSS form as elandtp159.sav. The survival experience of 68 patients enrolled in the Stanford Heart Transplantation Program (in days) are:

[d] p-Value derived from the chi-square statistic and its degrees of freedom using WinPEPI → WhatIs → P value.

0	1	1+	3	10	12	13+	14	15	23
25	26	29	30+	39	44	46	47	48	50
51	51	51	54	60	63	64	65	66	68
109+	127	136	147	161	166+	228	236+	253	280
297	304+	322	338+	338+	438+	455+	498+	551	588+
591+	624	659+	730	814+	836	837+	874+	897	994
1024	1105+	1263+	1350	1366+	1535+	1548+	1774+		

(A) Using the actuarial method, fill in the missing values in Table 17.10.
(B) Plot the cumulative survival function.
(C) Determine the approximate median survival time of the group. The median survival time is the point at which the cumulative survival proportion is 50%. This may be easiest to see on the graph plotted in part (B).

17.2 Use the data from the Tables 17.4 and 17.5 to determine the incidence proportion ratio associated with the treatment compared with the control group 8 years (96 months) into treatment.

17.3 Table 17.11 contains survival data from the University Group Diabetes Project (1970) for the group taking the tolbutamide (group 1) and the placebo group (group 0).
(A) Plot cumulative survival functions of both groups on a single axis. Interpret your plot.
(B) Calculate the Cochran–Mantel–Haenszel risk difference and a Mantel–Haenszel chi-square test statistic for these data. Interpret your results.

Table 17.10 Table shell for Exercise 17.1.

Interval start–end (days)	No. entering interval (N_i)	No. withdrawing during interval (W_i)	No. effectively exposed to risk (N'_i)	No. deaths during interval (A_i)	Proportion dying during interval ($\hat{q}_k$)	Proportion surviving during interval ($\hat{p}_k$)	Cumulative proportion surviving ($\hat{P}_k$)
0–49	68	3		16			
50–99	49	0		11			
100–199	38	2		4			
200–399	32	4		5			
400–699	23	6		2			
700–999	15	3		4			
1000–1299	8	2		1			
1300–1599	5	3		1			
1600+	1	1		0			

Table 17.11 Survival in tolbutamide and placebo groups, UGDP 1970.

Tolbutamide Year (k)	N_{1k}	W_{1k}	A_{1k}	N'_{1k}	$\hat{p}_{1k}$	$\hat{q}_{1k}$	$\hat{Q}_{1k}$
0–(1)	204	0	0	204	0.0000	1.0000	1.0000
1–(2)	204	0	5	204	0.0245	0.9755	0.9755
2–(3)	199	0	5	199	0.0251	0.9749	0.9510
3–(4)	194	5	5	191.5	0.0261	0.9739	0.9262
4–(5)	184	24	5	172	0.0291	0.9709	0.8992
5–(6)	155	41	4	134.5	0.0297	0.9703	0.8725
6–(7)	110	47	5	86.5	0.0578	0.9422	0.8221
7–(8)	58	33	1	41.5	0.0241	0.9759	0.8022
Placebo Year (k)	N_{0k}	W_{0k}	A_{0k}	N'_{0k}	$\hat{p}_{0k}$	$\hat{q}_{0k}$	$\hat{Q}_{0k}$
0–(1)	205	0	0	205	0.0000	1.0000	1.0000
1–(2)	205	0	5	205	0.0244	0.9756	0.9756
2–(3)	200	0	4	200	0.0200	0.9800	0.9561
3–(4)	196	4	4	194	0.0206	0.9794	0.9364
4–(5)	188	23	4	176.5	0.0227	0.9773	0.9152
5–(6)	161	43	3	139.5	0.0215	0.9785	0.8955
6–(7)	115	50	1	90	0.0111	0.9889	0.8855
7–(8)	64	36	0	46	0.0000	1.0000	0.8855

Source: Elandt-Johnson and Johnson (1980, p. 250).

Table 17.12 Solution to Exercise 17.1, Part A.

Interval start–end (Days)	No. entering interval (N_k)	No. withdrawing during interval (W_k)	No. effectively exposed to risk (N'_k)	No. deaths during interval (A_k)	Proportion dying during interval ($\hat{q}_k$)	Proportion surviving during interval ($\hat{p}_k$)	Cumulative proportion surviving at end[a] ($\hat{P}_k$)
0–49	68	3	66.5	16	0.2406	0.7594	0.7594
50–99	49	0	49	11	0.2245	0.7755	0.5889
100–199	38	2	37	4	0.1081	0.8919	0.5253
200–399	32	4	30	5	0.1667	0.8333	0.4377
400–699	23	6	20	2	0.1000	0.9000	0.3939
700–999	15	3	13.5	4	0.2963	0.7037	0.2772
1000–1299	8	2	7	1	0.1429	0.8571	0.2396
1300–1599	5	3	3.5	1	0.2857	0.7143	0.1697
1600+	1	1	0.5	0	0.0000	1.00000	—

[a] Elandt-Johnson and Johnson (1980) reports cumulative percentage surviving to *beginning* of interval. This table reports cumulative percentages at the end of the interval.

References

Abramson, J.H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations*, **8** (1), 1 *http://www.epi-perspectives.com/content/8/1/1*.

Cochran, W.G. (1954) Some methods for strengthening the common chi-square tests. *Biometrics*, **10**, 417–451.

Cox, D.R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, **34**, 187–220.

Crowley, J., and Hu, M. (1974) Covariance Analysis of Heart Transplant Survival Data. Stanford University, Technical Report No. **2**, pp. 1–35.

Elandt-Johnson, R.C., and Johnson, N.L. (1980) *Survival Models and Data Analysis*, John Wiley & Sons, Inc., New York.

Kaplan, E.L., and Meier, P. (1958) Non-parametric estimation for incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Mantel, N. (1963) Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, **58**, 690–700.

Mantel, N. (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, **50**, 163–170.

UGDP (1970) University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. *Diabetes*, **19** (Suppl. 2).

CHAPTER 18
Current Life Tables

18.1 Introduction

18.2 Complete life table

- Predicting probabilities from rates
- Special circumstances surrounding the first year of life
- General formula
- Constructing a complete life table

18.3 Abridged life table

Exercises

References

> By this simple and elegant table the mean duration of human life, uncertain as it appears to be, and as reference to individuals, can be determined with the greatest accuracy in nations, or in still smaller communities.
>
> *Farr (1885, p. 450)*

18.1 Introduction

Generational (cohort) life tables versus current ("cross-sectional") life tables. A **life table** is a scheme for expressing mortality over an entire lifetime. To accomplish this, we could in theory apply either the actuarial or Kaplan–Meier techniques presented in the previous chapter to the experience of an entire generation, from birth of the first cohort member until the death of the last person in the cohort. This would construct a **generational (cohort) life table**. However, compilation of just a single generational life table would require more than a century's worth of data, which is not feasible in most instances and is not useful for predicting current life expectancies. To work around these constraints, we may use the current mortality experience of a population to construct a life table for a *hypothetical* cohort. Life tables constructed in this manner are called **current ("cross-sectional") life tables**.

US life table survival curves. Figure 18.1 shows current life table survival curves for the US 1900, 1950, and 1997 populations. These curves reveal large improvements in longevity over the span of the 20th century. Between 1900 and 1950, the most dramatic improvements in survival were seen at younger ages. Between 1950 and 1997, improvements were most notable at older ages. The effect

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

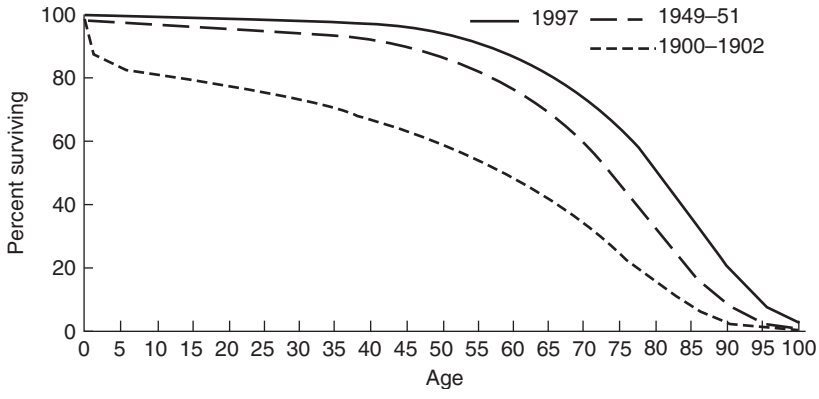


Figure 18.1 Percent surviving, United States, 1900–02, 1949–51, and 1997. Source: Anderson (1999).

has been to **rectangularize** the survival curve over time. This, combined with lower fertility rates, has resulted in the demographic and epidemiologic transitions discussed in Section 1.3.

Current life tables are not longitudinal analyses. Current life tables are based on an examination of current mortality rates. Although examinations of current rates in open populations may appear similar to cohort rates, they are quite different from each other. Examination of current rates in open populations does not require the follow-up of individuals over time and, therefore, does not represent a true survival analysis.

Important principles conveyed through notation. As you read this chapter you will encounter many formulas and notational conventions that are necessary when describing population mortality. You should not rush through these formulas or view them strictly as computational devices, for many important epidemiologic principles are inherent within them.

18.2 Complete life table

Stationary population model. In constructing a current life table, it is necessary to assume the population is stationary. A **stationary population** is a population in which:

1 There are a given number of births each year.
2 Births are uniformly distributed over the year.
3 The mortality experience of the population is distributed in such a way as to balance births.

(Elandt-Johnson and Johnson, 1980).

This creates a population *model* of constant size and age structure, allowing us to use the mortality experience of the current population to describe the survival of a *hypothetical* cohort.

Complete vs. abridged life tables. Current life tables can be constructed to account for single years of life or for any other age interval (e.g., for every 5 years of life). Life tables that are constructed for single years of age are **complete life tables**, while those constructed for 5-year age intervals are **abridged life tables**. Of course, the terms ''complete'' and ''abridged'' are arbitrary since a life table computed for monthly age intervals would be more complete than a complete life table, and a life table computed for 10-year age intervals would be more abridged than a quinquennial abridged life table. Nevertheless, the ''complete'' and ''abridged'' terminology is well established.

Predicting probabilities from rates

Let us start by constructing a complete life table. The most basic type of information needed to construct a complete life table is the probability describing the likelihood of death within each year of life. **Age intervals** are denoted $[x, x + 1)$, meaning age x to just before age $x + 1$. Sometimes this is written age x lbd, meaning age x as of the last birthday.

Probabilities of death are denoted ${}_{x+1}p_x$, or, when working with a single year of life, p_x (the left-sided subscript is unnecessary when working with single-year intervals). For example, p_{30} denotes the probability of death between 30 and 31 years of age. Probabilities of death are not observed directly, but can be estimated from observed death rates.

Comment. Recall that we are using p to denote probabilities of death. Other sources may refer to this with the symbol q (the complement of p) or R (for ''risk'').

Let λ_x denote the **death rate** for age x lbd. For example, λ_{30} denotes the death rate for 30-year olds. It is important to recognize that death probabilities *(p_x)* differ fundamentally from death rates (λ_x). The former represents incidence proportions *(risks)*, while the latter represents incidence *rates* (see Section 6.1). Nevertheless, by assuming a stationary population model, we derive p_x from λ_x as follows. Let:

$A_x \equiv$ number of deaths in age $[x, x + 1)$
$\overline{N}_x \equiv$ ''the population existing'' at age x at a specific time
$\Delta t \equiv$ length of the observation period (e.g., 3 years of data).

The **annual mortality rate** for age x is

$$\lambda_x = \frac{A_x}{\overline{N}_x \cdot \Delta t} \tag{18.1}$$

For example, if we observed 14 610 deaths in 30-year olds between 1 January 1996, and 31 December 1998, and the census estimate of the population size reveals 4 272 274 30-year-olds, then $\lambda = 14\,610/(4\,272\,241 \times 3 \text{ years}) = 0.001\,14 \text{ year}^{-1}$.

Now consider a hypothetical cohort of N_x people turning age x. At exact age $x + 1$ there are $N_{x+1} = N_x - A_x$ people. If we assume deaths occur uniformly in this age group over the course of the year,[a] the size of the group that is age x_{lbd} group diminishes linearly, and the number of people existing (alive) at midyear is $\overline{N}_x = N_x - \frac{1}{2}A_x$ (Figure 18.2).

[a] The assumption is a good one for all age groups except 0–1-year olds and, perhaps, the very old.

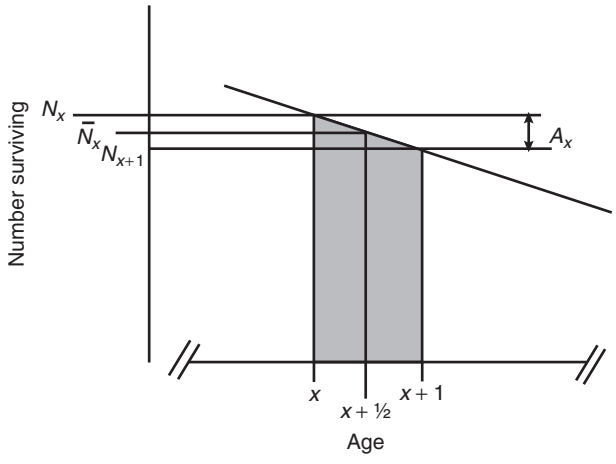


Figure 18.2 Number surviving over single year of age [x, $x + 1$); assumes deaths occur uniformly over the year.

Thus:

$$N_x = \overline{N}_x + \left(\frac{1}{2}\right) A_x \tag{18.2}$$

and the **predicted probability of death during age [x, $x + 1$)** is

$$\hat{p}_x = \frac{A_x}{N_x} = \frac{A_x}{\overline{N}_x + (1/2)\, A_x} \tag{18.3}$$

Since $\lambda_x = A_x/\overline{N}$, $A_x = \lambda_x \cdot \overline{N}_x$ and

$$\hat{p}_x = \frac{A_x}{\overline{N}_x + (1/2)\, A_x} = \frac{\lambda_x \cdot \overline{N}_x}{\overline{N}_x + (1/2)\left(\lambda_x \cdot \overline{N}_x\right)} = \frac{\lambda_x}{1 + (1/2)\lambda_x} \tag{18.4}$$

For example, when $\lambda_{30} = 0.001\,14$

$$\hat{p}_x = \frac{0.00114}{1 + (1/12)\,(0.00114)} = 0.001139$$

Special circumstances surrounding the first year of life

The above set of formulas assumes a linear decline in the size of the x-year old population over the year. This is a good assumption at most ages, but does not hold for [0, 1) where mortality is greatest soon after birth. Based on empirical research, the expected fraction of the year of life lived in those who die in the first year of life is 0.1–0.2 (closer to 0.1). For example, the expected fraction of the year of life lived in the United States in 1997 for those who died before reaching age 1 was 0.13 (Anderson, 1999). By using this fraction, we derive the number of live births based on the number of 0–1-year olds existing (alive) according to the formula $N_0 = \overline{N}_x + (1 - 0.13)A_0$. Therefore, for 0–1-year olds,

$$\hat{p}_0 = \frac{A_0}{\overline{N}_0 + (1 - 0.13)A_0} = \frac{\lambda_0 \overline{N}_0}{\overline{N}_0 + (1 - 0.13)\left(\lambda_0 \overline{N}_0\right)} = \frac{\lambda_0}{1 + (1 - 0.13)\lambda_0} \tag{18.5}$$

General formula

Let f_x represent the **expected fraction of the [x, x + 1) age interval that is lived among those that die during that interval**. The predicted probability of death during age x_{lbd} is

$$\hat{p}_x = \frac{\lambda_x}{1 + (1 - f_x)\,\lambda_x} \tag{18.6}$$

where $f_x = 0.5$ for all ages *except* 0–1. Assume $f_0 \cong 0.13$ in industrialized societies.

Comment. Modern life tables include an additional adjustment for the elderly, which is not covered here.

Constructing a complete life table

Once values for p_x are calculated, the remaining quantities in the life table are easy to compute. In the interest of maintaining consistency with established life table notation, the following symbols are adopted:

p_x denotes the probability of death during age [x, x + 1). This quantity is computed with Formula (18.6). (This probability is denoted q_x in government life tables.)

l_x denotes the expected number of individuals *living* to exact age x in a hypothetical cohort of 100 000 newborns. By convention, l_0 is set to 100 000. The number surviving to age 1 in the hypothetical cohort is denoted l_1; the number surviving to age 2 is l_2, and so on. The expected number remaining alive to age x + 1 is calculated as:

$$l_{x+1} = l_x - d_x \tag{18.7}$$

where

d_x denotes the *expected* number of *deaths* in the hypothetical cohort between ages x and x + 1, calculated as the product of l_x and p_x:

$$d_x = l_x p_x \tag{18.8}$$

L_x denotes the *expected* number of *person-years lived* in the hypothetical cohort from age x to x + 1. Each member of the hypothetical cohort who survives the year contributes one person-year to L_x. Each person who dies in the interval [x, x + 1) contributes fraction f_x of a person-year to L_x. (As mentioned, $f_x = 0.5$ for all ages other than age 0, where $f_0 \cong 0.13$ in the US population.) The person-years lived in the hypothetical cohort in age [x, x + 1) is

$$L_x = l_x - (1 - f_x)\,d_x \tag{18.9}$$

The expected number of person-years lived in the oldest age group is

$$L_{100+} = \frac{l_{100+}}{\lambda_{100+}} \tag{18.10}$$

T_x denotes the sum *total* of person-years lived by the hypothetical cohort from age x onward. This is equal to the sum of the L_x values from age x onward:

$$T_x = L_x + L_{x+1} + \cdots + L_{100+} \tag{18.11}$$

e_x denotes the number of years *expected* to be lived by a person at exact age x and is equal to the total number of person-years remaining in the hypothetical cohort divided by the number of people living to age x:

$$e_x = \frac{T_x}{l_x} \tag{18.12}$$

Illustrative Example 18.1 1997 US life table

Table 18.1 is a reproduction of the complete current life table for the 1997 US population. In this table, the probability of death between 0 and 1, p_0 was determined to be 0.007 23. Therefore, $d_0 = (l_0)(p_0) = (100\,000)(0.007\,23) = 723$. The expected number remaining alive to age 1, $l_1 = l_0 - d_0 = 100\,000 - 723 = 99\,277$. The number of person-years in the 0–1-year old range is $L_0 = l_0 - (1 - f_0)(d_0) = 100\,000 - (1 - 0.13)(723) = 99\,371$ person-years. The cohort is assumed to have 1559 centenarians ($\lambda_{100+} = 1559$), and the death rate in centenarians is assumed to be 0.4017 per year ($l_{100+} = 0.4017$). Therefore, $L_{100+} = 1559/0.417 = 3871$. The number of person-years remaining in, say, 30-year olds is $T_{30} = 97\,518 + 97\,405 + \cdots + 3871 = 4\,689\,680$. Therefore, the expected number of years of life remaining at exact age 30, $e_{30} = 4\,689\,680/97\,574 = 48.1$ years (i.e., a 30-year old is expected to die at age 30 + 48.1 = 78.1).

Table 18.1 Current life table for total population: United States, 1997.

Age	Proportion dying during age interval q_x	Number living at beginning of age interval l_x	Number dying during age interval d_x	Stationary population in age interval L_x	Stationary population in this and all subsequent age intervals T_x	Life expectancy at beginning of age Interval e_x
0–1	0.007 23	100 000	723	99, 371	7 650 789	76.5
1–2	0.000 55	99 277	55	99 250	7 551 418	76.1
2–3	0.000 36	99 223	36	99 205	7 452 168	75.1
3–4	0.000 29	99 187	29	99 172	7 352 963	74.1
4–5	0.000 23	99 158	23	99 146	7 253 791	73.2
5–6	0.000 21	99 135	21	99 125	7 154 644	72.2
6–7	0.000 20	99 114	20	99 104	7 055 520	71.2
7–8	0.000 19	99 094	19	99 085	6 956 416	70.2
8–9	0.000 17	99 076	17	99 067	6 857 330	69.2
9–10	0.000 15	99 059	15	99 051	6 758 263	68.2
10–11	0.000 14	99 043	14	99 037	6 659 212	67.2
11–12	0.000 14	99 030	14	99 023	6 560 175	66.2
12–13	0.000 19	99 016	19	99 006	6 461 153	65.3
13–14	0.000 28	98 997	28	98 983	6 362 147	64.3
14–15	0.000 41	98 969	40	98 949	6 263 164	63.3
15–16	0.00055	98 929	54	98 901	6 164 215	62.3
16–17	0.00068	98 874	67	98 841	6 065,313	61.3
17–18	0.000 78	98 807	77	98 768	5 966 473	60.4
18–19	0.000 85	98 730	84	98 688	5 867 704	59.4
19–20	0.000 89	98 646	88	98 602	5 769 016	58.5
20–21	0.000 93	98 558	92	98 512	5 670 414	57.5
21–22	0.000 98	98 467	96	98 418	5 571 902	56.6
22–23	0.001 01	98 370	99	98 321	5 473 483	55.6
23–24	0.001 01	98 272	100	98 222	5 375 162	54.7

Table 18.1 (*continued*)

Age	Proportion dying during age interval q_x	Number living at beginning of age interval l_x	Number dying during age interval d_x	Stationary population in age interval L_x	Stationary population in this and all subsequent age intervals T_x	Life expectancy at beginning of age Interval e_x
24–25	0.001 01	98 172	99	98 123	5 276 940	53.8
25–26	0.001 00	98 073	98	98 024	5 178 818	52.8
26–27	0.000 99	97 975	97	97 927	5 080 794	51.9
27–28	0.001 00	97 878	98	97 829	4 982 867	50.9
28–29	0.001 03	97 780	101	97 730	4 885 037	50.0
29–30	0.001 08	97 679	106	97 627	4 787 307	49.0
30–31	0.001 14	97 574	111	97 518	4 689 680	48.1
31–32	0.001 19	97 463	116	97 405	4 592 162	47.1
32–33	0.001 26	97 347	122	97 286	4 494 757	46.2
33–34	0.001 33	97 225	129	97 160	4 397 471	45.2
34–35	0.001 40	97 096	136	97 027	4 300 311	44.3
35–36	0.00149	96 959	144	96 887	4 203 284	43.4
36–37	0.001 57	96 815	152	96 739	4 106 396	42.4
37–38	0.00167	96 663	161	96 582	4 009 657	41.5
38–39	0.001 78	96 502	172	96 416	3 913 075	40.5
39–40	0.001 92	96 330	185	96 237	3 816 659	39.6
40–41	0.002 06	96 145	198	96 046	3 720 422	38.7
41–42	0.002 22	95 947	213	95 841	3 624 376	37.8
42–43	0.002 39	95 734	229	95 620	3 528 535	36.9
43–44	0.002 57	95 506	246	95 383	3 432 915	35.9
44–45	0.00278	95 260	264	95 128	3 337 532	35.0
45–46	0.003 00	94 996	285	94 853	3 242 404	34.1
46–47	0.00325	94 710	308	94 556	3 147 551	33.2
47–48	0.003 52	94 402	332	94 236	3 052 995	32.3
48–49	0.003 80	94 070	358	93 891	2 958 759	31.5
49–50	0.004 11	93 712	385	93 519	2 864 868	30.6
50–51	0.004 44	93 327	415	93 120	2 771 349	29.7
51–52	0.004 82	92 912	448	92 688	2 678 229	28.8
52–53	0.005 24	92 464	485	92 221	2 585 541	28.0
53–54	0.005 71	91 979	525	91 717	2 493 320	27.1
54–55	0.006 23	91 454	570	91 169	2 401 603	26.3
55–56	0.006 85	90 884	622	90 573	2 310 434	25.4
56–57	0.007 55	90 262	681	89 921	2 219 861	24.5
57–58	0.008 33	89 580	746	89 208	2 129 940	23.8
58–59	0.009 16	88 835	814	88 428	2 040 733	23.0
59–60	0.010 05	88 021	884	87 579	1 952 305	22.2
60–61	0.011 01	87 136	959	86 657	1 864 727	21.4
61–62	0.012 08	86 177	1041	85 657	1 778 070	20.6
62–63	0.013 21	85 136	1125	84 574	1 692 413	19.9
63–64	0.014 39	84 011	1209	83 407	1 607 839	19.1
64–65	0.015 60	82 802	1292	82 156	1 524 433	18.4
65–66	0.016 79	81 510	1368	80 826	1 442 277	17.7
66–67	0.018 02	80 142	1444	79 419	1 351 451	17.0
67–68	0.019 48	78 697	1533	77 931	1 282 032	16.3
68–69	0.021 27	77 154	1642	76 343	1 204 101	15.6
69–70	0.023 38	75 522	1765	74 640	1 127 758	14.9
70–71	0.025 65	73 757	1892	72 811	1 053 118	14.3

Table 18.1 (*continued*)

Age	Proportion dying during age interval q_x	Number living at beginning of age interval l_x	Number dying during age interval d_x	Stationary population in age interval L_x	Stationary population in this and all subsequent age intervals T_x	Life expectancy at beginning of age Interval e_x
71–72	0.027 99	71 865	2011	70 859	980 307	13.6
72–73	0.030 43	69 854	2126	68 791	909 447	13.0
73–74	0.032 97	67 728	2233	66 612	840 657	12.4
74–75	0.035 63	65 495	2334	64 328	774 045	11.8
75–76	0.038 43	63 162	2427	61 948	709 716	11.2
76–77	0.041 47	60 735	2519	59 475	647 768	10.7
77–78	0.044 94	58 216	2616	56 908	588 293	10.1
78–79	0.049 04	55 600	2726	54 237	531 385	9.6
79–80	0.053 85	52 874	2847	51 450	477 148	9.0
80–81	0.059 38	50 026	2971	48 541	425 698	8.5
81–82	0.055 55	47 055	3084	45 613	377 158	8.0
82–83	0.072 41	43 971	3184	42 379	331 644	7.5
83–84	0.079 90	40 787	3259	39 158	289 265	7.1
84–85	0.088 12	37 528	3307	35 875	250 107	6.7
85–86	0.096 53	34 221	3303	32 570	214 232	6.3
86–87	0.105 56	30 918	3264	29 286	181 663	5.9
87–88	0.115 39	27 654	3191	26 059	152 376	5.5
88–89	0.126 16	24 463	3086	22 920	126 318	5.2
89–90	0.138 02	21 377	2950	19 902	103 398	4.8
90–91	0.150 85	18 427	2780	17 037	83 496	4.5.
91–92	0.164 29	15 647	2571	14 362	66 459	4.2.
92–93	0.17813	13 076	2329	11 912	52 097	4.0
93–94	0.192 50	10 747	2069	9713	40 186	3.7
94–95	0.207 64	8678	1802	7777	30 473	3.5
95–96	0.223 54	6876	1537	6108	22 696	3.3
96–97	0.239 99	5339	1281	4699	16 588	3.1
97–98	0.256 53	4058	1041	3537	11 889	2.9
98–99	0.272 95	3017	823	2605	8352	2.8
99–100	0.289 15	2193	634	1876	5747	2.6
100+	1.000 00	1559	1559	3871	3871	2.5

Source: Anderson (1999).

18.3 Abridged life table

Table 18.2 is an abridged life table for the 1998 US population. Construction and interpretation of abridged life tables parallel that of complete current life tables with one exception: the abridged life table lists probabilities of death at 5-year age intervals for all age ranges except for from 0 to 1 and from 1 to 4. The **annual mortality rate for ages [x, $x + n$)** is calculated based on vital statistics over an observation period of length Δt:

$$ {}_n\lambda_x = \frac{\text{no. of deaths in}[x,\ n+x)\ \text{ over period }\ \Delta t}{\text{population size at time } t' \text{ for age group}[x,\ n+x)\ \times\ \Delta t} $$

Table 18.2 Abridged life table for total population: United States, 1998.

Age interval	Proportion dying	Of 100 000 born alive		Stationary population		Average remaining lifetime
Period of life between two exact ages stated in years (1) x to $x + n$	Proportion of persons alive at beginning of age interval dying during interval (2) ${}_nq_x$	Number living at beginning of age interval (3) l_x	Number dying during age interval (4) ${}_nd_x$	In the age interval (5) ${}_nL_x$	In this and all subsequent age interval (6) T_x	Average number of years of life remaining at beginning of age interval (7) e_x
0–1	0.007 21	100 000	721	99 370	7 671 400	76.7
1–5	0.001 39	99 279	138	396 785	7 572 030	76.3
5–10	0.000 89	99 141	88	495 473	7 175 244	72.4
10–15	0.001 10	99 053	109	495 057	6 679 771	67.4
15–20	0.003 53	98 944	349	493 926	6 184 714	62.5
20–25	0.004 76	98 595	469	491 820	5 690 788	57.7
25–30	0.004 87	98 126	478	489 450	5 198 958	53.0
30–35	0.006 00	97 648	586	486 840	4 709 518	48.2
35–40	0.008 19	97 062	795	483 428	4 222 678	43.5
40–45	0.011 76	96 267	1132	478 870	3 739 250	38.8
45–50	0.017 28	95 135	1644	471 811	3 260 580	34.3
50–55	0.025 64	93 491	2397	461 839	2 788 769	29.8
55–60	0.040 09	91 094	3652	446 966	2 326 930	25.5
60–65	0.063 02	87 442	5511	424 280	1 879 964	21.5
65–70	0.094 37	81 931	7732	391 364	1 455 684	17.8
70–75	0.142 39	74 199	10 565	345 660	1 064 320	14.3
75–80	0.206 04	63 534	13 111	286 484	718 660	11.3
80–85	0.316 41	50 523	15 986	213 528	432 176	8.6
85–90	0.461 04	34 537	15 923	131 897	218 650	6.3
90–95	0.815 02	18 614	11 448	62 020	86 753	4.7
95–100	0.754 26	7166	5405	20 150	24 733	3.5
100 and over	1.000 00	1761	1761	4583	4583	2.6

Source: Murphy (2000).

Table 18.3 Number of survivors of 100 000 live births, US white females.

Age	1997	1959–1961	1919–1921	1900–1902
0	100 000	100 000	100 000	100 000
1	99 464	98 036	93 608	88 939
5	99 353	97 709	90 721	83 426
10	99 283	97 525	89 564	81 723
15	99 198	97 375	88 712	80 680
20	98 986	97 135	87 281	78 978
25	98 764	96 844	85 163	76 588
30	98 516	96 499	82 740	73 887
35	98 180	96 026	80 206	70 971
40	97 702	95 326	77 624	67 935
45	96 993	94 228	74 871	64 677
50	95 922	92 522	71 547	61 005
55	94 193	89 967	67 323	56 509
60	91 412	86 339	61 704	50 752
65	87 106	80 739	54 299	43 806
70	80 905	72 507	44 638	35 206
75	71 921	60 461	32 777	25 362
80	59 627	44 676	20 492	15 349
85	43 261	26 046	9909	7149
90	24 704	10 219	3372	2322
95	9603	2203	721	448
100	2178	265	63	41

Source: Anderson (1999, Table 10, p. 26).

The predicted probability of death in $[x, x + n)$ is

$$ {}_n\hat{p}_x = \frac{{}_n\lambda_x}{1/n\left[1 + n\left(1 - {}_nf_x\right){}_n\lambda_x\right]} \tag{18.14} $$

where ${}_nf_x$ is the expected proportion of the $[x, x + n)$ interval lived for those aged x who die in $[x, x + n)$.

Once values for ${}_np_x$ are estimated, the life table is constructed with the following formulas:

$$ l_{x+n} = l_x\left(1 - {}_np_x\right) \tag{18.15} $$

$$ {}_nd_x = l_x - l_{x+n} = l_x({}_np_x) \tag{18.16} $$

$$ {}_nL_x = n\left[lx - \left(1 - {}_nf_x\right){}_nd_x\right] \tag{18.17} $$

for all age groups except the last, in which case

$$ L_{100+} = \frac{l_{100+}}{\lambda_{100+}} \tag{18.18} $$

$$ T_x = {}_nL_x + {}_nL_{x+n} + \cdots + {}_\infty L_{100} \tag{18.19} $$

and

$$ e_x = \frac{T_x}{l_x} \tag{18.20} $$

Table 18.4 Shell for answering Exercise 18.2. Abridged life table, US white women, 1919–1921.

Age range x to $x+n$	Proportion dying (1) ${}_nq_x$	No. living at beginning of age interval (2) l_x	No. dying during the age interval (3) ${}_nd_x$	Expected person-years in age interval (4) ${}_nL_x$	Expected person-years in this and all subsequent age intervals (5) T_x	Expected no. of years of life remaining as of age x (6) e_x
0–1		100 000				
1–5		93 608				
5–10		90 721				
10–15		89 564				
15–20		88 712				
20–25		87 281				
25–30		85 163				
30–35		82 740				
35–40		80 206				
40–45		77 624				
45–50		74 871				
50–55		71 47				
55–60		67 323				
60–65		61 704				
65–70		54 299				
70–75		44 638				
75–80		32 777				
80–85		20 492				
85–90		9909				
90–95		3372				
95–100		721				
100+		63				

Exercises

18.1 Table 18.3 lists the expected number of survivors out of 100 000 US white females born alive for the periods 1997, 1959–1961, 1919–1921, and 1900–1902. (Anderson, 1999, Table 10, p. 26). Plot curves for these four populations (see Figure 18.1 for an example). What does it mean when we say the survival curve has become rectangularized?

18.2 Using the data in Table 18.3, construct an abridged current life table for US white women for 1919–1921. A shell for the table is provided as Table 18.4. Start by filling in column (3) (${}_nd_x = l_x - l_{x+n}$). Then, calculate column (1) (${}_nP_x = nd_x \div l_x$). To calculate the values for column (4), assume ${}_1f_0 = 0.1$, ${}_4f_1 = 0.2$, and ${}_nf_x = 0.5$ for all other ages. Recall that ${}_nL_x = n[l_x - (1 - {}_nf_x)\,{}_nd_x]$ except in the oldest age category, where $L_{100+} = l_{100+}/\lambda_{100+}$. Assume $\lambda_{100} = 0.4$ year^{-1}. (Numerical assumptions have been rounded to simplify computations.)

References

Anderson, R.N. (1999) United States life tables, 1997. *National Vital Statistics Reports*, **47**, 1–40.

Elandt-Johnson, R.C., and Johnson, N.L. (1980) *Survival Models and Data Analysis*, John Wiley & Sons, Inc., New York.

Farr, W. (1885) *Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr. (1975 Reprint)*, Offices of the Sanitary Institute, Metuchen, NJ.

Murphy, S.L. (2000) Deaths: Final data for 1998. *National Vital Statics Reports*, **48**, 1–105.

CHAPTER 19

Random Distribution of Cases in Time and Space

19.1 Introduction

19.2 The Poisson distribution

- Use of the Poisson formula
- Calculating the expected number of cases
- *Post hoc* identification of clusters

19.3 Goodness of fit of the Poisson distribution

- Fitting the Poisson distribution

19.4 Summary

Exercises

References

The relations between probability and experience are also still in need of clarification.

Karl Popper (1959, p. 133)

19.1 Introduction

Random distribution in time and space Epidemiologists are often called upon to evaluate whether an observed number of cases in a population is greater than expected and whether any such increase is beyond what could be expected due to chance. Random fluctuations in the number of cases in a population are to be expected, and are easily addressed by statistical analyses.

Don't be fooled by randomness Figure 19.1 is a computer-generated image of 50 dots randomly distributed on a grid of 50 squares. The average number of dots per box is 1 with the following distribution: 18 of the squares contain no dots, 19 contain 1 dot, 9 contain 2 dots, 3 contain 3 dots, and 1 contains 4 dots. Now imagine that each square on this grid represents a community, and each dot represents a case of some relatively rare disease. (The communities are of equal size and age distribution.) If we were to select the community with the 4 cases, we would be correct in saying that the occurrence was 4 times the expected rate. However, this is merely an artifact of the random distribution of cases. Therefore, in assessing whether an occurrence is greater than expected, we must factor in expected geographic and temporal random

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

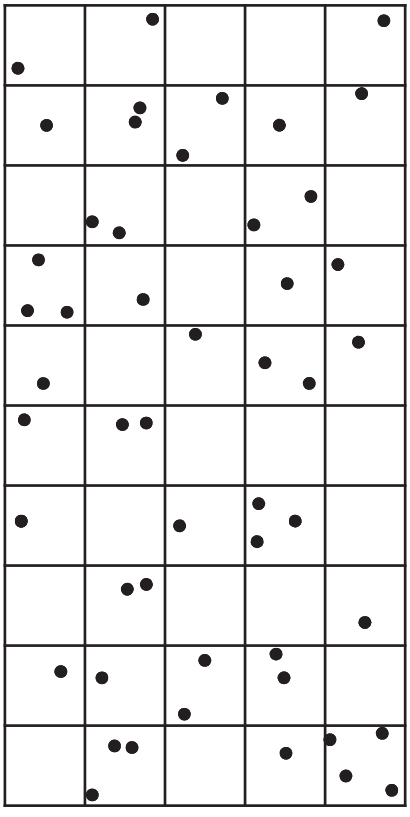


Figure 19.1 Random distribution of 50 dots on a grid with 50 squares.

fluctuations. The question then becomes: At what point can we declare a distribution of cases to be nonrandom?

19.2 The Poisson distribution

The Poisson distribution is a probability distribution that is well suited for describing the random occurrence of rare events over time. The Poisson formula is

$$\Pr(A = a) = \frac{e^{-\mu}\mu^{a}}{a!} \tag{19.1}$$

Where:

$A \equiv$ the variable number of cases over a given amount of person-time
$\Pr(A = a) \equiv$ the probability of observing exactly a cases
$e \equiv$ the universal constant that forms the base of natural logarithms ($e = 2.718\,281$)
$\mu \equiv$ the expected number of cases in the population
$a! \equiv$ the mathematical operation "a factorial" = $a(a - 1)(a - 2)\ \ldots\ (1)$. For example, $4! = (4)(3)(2)(1) = 24$. By definition, $0! = 1$.

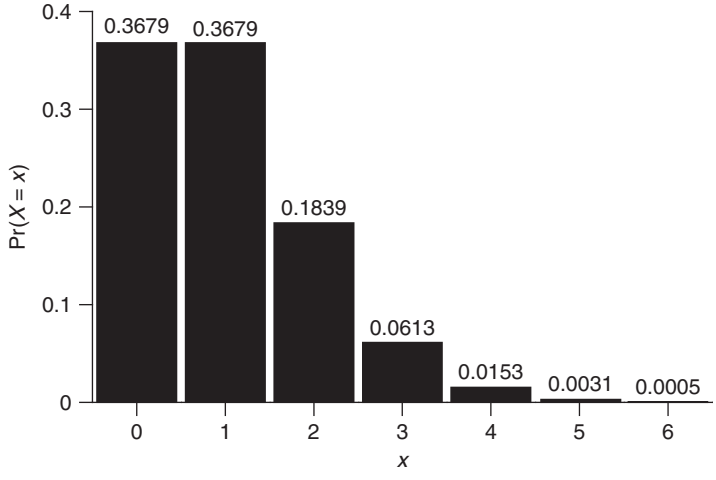


Figure 19.2 Poisson distribution, $\mu = 1$.

Use of the Poisson formula

To illustrate use of the Poisson formula, let us consider the probability of observing no cases ($A = 0$) in a population in which one case is expected ($\mu = 1$). Accordingly,

$$\Pr(A = 0) = \frac{e^{-1}1^0}{0!} = 0.3679$$

The probability of observing one case when one case is expected is

$$\Pr(A = 1) = \frac{e^{-1}1^1}{1!} = 0.3679$$

The probability of observing two cases is

$$\Pr(A = 2) = \frac{e^{-1}1^2}{2!} = 0.1839$$

The probability of observing three cases is

$$\Pr(A = 3) = \frac{e^{-1}1^3}{3!} = 0.0613$$

The probability of observing four cases is

$$\Pr(A = 4) = \frac{e^{-1}1^4}{4!} = 0.0153$$

and so on. Figure 19.2 displays a bar chart of this distribution.

Calculating the expected number of cases

Use of the Poisson formula requires knowledge of the expected number of cases in a population. In epidemiologic investigations, this information comes from knowledge

of the population size (n) and expected rate (λ_0):

$$\mu = (n)\left(\lambda_0\right) \tag{19.2}$$

Illustrative Example 19.1 Poisson distribution (rare cancer)

Suppose a rare form of cancer has an expected rate of 1 per million person-years in a population of a given age distribution. In a city of, say, 100 000 (with the given age distribution), we expect to see $\mu = (100\,000)(1 \times 10^{-6}$ year) = 0.1 cases per year. The Poisson distribution for $\mu = 0.1$ is calculated:

$$\Pr(A = 0) = \frac{\left(e^{-0.1}\right)\left(0.1^0\right)}{0!} = \frac{(0.9048)\,(1)}{(1)} = 0.9048$$

$$\Pr(A = 1) = \frac{\left(e^{-0.1}\right)\left(0.1^1\right)}{1!} = \frac{(0.9048)\,(0.01)}{(1)} = 0.0905$$

$$\Pr(A = 2) = \frac{\left(e^{-0.1}\right)\left(0.1^2\right)}{2!} = \frac{(0.9048)\,(0.01)}{(2)} = 0.0045$$

$$\Pr(A = 3) = \frac{\left(e^{-0.1}\right)\left(0.1^3\right)}{3!} = \frac{(0.9048)\,(0.001)}{(6)} = 0.000\,15$$

$$\Pr(A = 4) = \frac{\left(e^{-0.1}\right)\left(0.1^4\right)}{4!} = \frac{(0.9048)\,(0.0001)}{(24)} = 0.000\,003\,8$$

(and so on).

Figure 19.3 displays these probabilities in the form of a bar chart. From this distribution we can say that the city will experience zero cases most years (90.5% of the time), one case 9.0% of the time, and two *or more* cases about 0.5% of the time (the right tail of the distribution). We may therefore say that under this model, we will see two or more cases every 200 years (on the average).

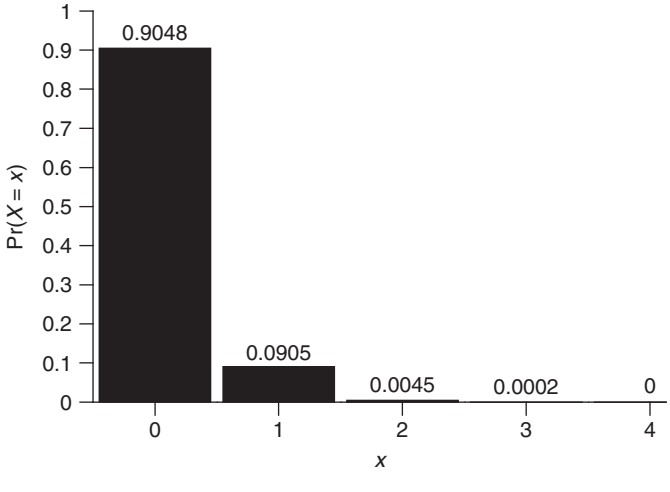


Figure 19.3 Poisson distribution for Illustrative Example 19.1, $\mu = 0.1$.

Epidemiologic computing WinPEPI[a] (Abramson, 2011) will compare an observed number of cases to an expected number of cases under a Poisson model. Use WinPEPI → Describe → H. Compare SMR or indirectly standardized rate. Look for the one-sided Fisher's exact results to replicate the results in Illustrative Example 19.1.

Post hoc identification of clusters

The term *clustering* in epidemiology is usually reserved for describing unusually high accumulations of rare diseases in a circumscribed time and space. By understanding the Poisson distribution as a description of random occurrences in time and space, we can appreciate that a certain amount of clustering is to be expected—*somewhere* there will be a clustering of cases.

In 1989, state health departments in the United States received approximately 1500 requests to investigate cancer clusters (Greenberg and Wartenberg, 1991). Many of these requests for investigations turned out to be normal occurrences or artifacts of inflated reporting. Those that did represent true increases in occurrence were often difficult to evaluate. Therefore, in 1989, the Centers for Disease Control and Prevention convened a national conference to discuss the study of cancer clusters. The conference clarified the following difficulties surrounding such investigations (Rothman, 1990):

- Perceived clusters often include different types of cancers, thus reducing the likelihood that they resulted from a common exposure.
- Many reported clusters include too few cases to reach reliable statistical conclusions.
- Regional boundaries of clusters are rarely demarcated, making it difficult to determine the size of the population at risk that gave rise to the cases in question.
- Regional boundaries of clusters may have been arbitrarily altered to make the cluster seem more substantial or inclusive.
- Conclusions about the perceived clusters may not be reliable because of differences in the sensitivity of statistical mapping techniques used for their detection.
- Causal exposures are often unspecified and, when specified, are often insufficiently intense to explain the perceived cluster.
- Chance can never really be ruled out as an ''after the fact'' explanation for a cluster—even when the statistical chances of an observation are small, rare events are inevitable if enough possibilities are considered.

Despite difficulties encountered in studying clusters, most public health agencies agree that it is good public relations to respond to community concerns about cancer clusters. If a true increase in cancer frequency does exist, citizens can take appropriate action. If a true increase in cancer frequency does *not* exist, the worries of citizens can be alleviated.[b] Cluster investigations also provide the opportunity for public health agencies to demonstrate their responsiveness to public concerns and to educate the public (Bender *et al*., 1990). Therefore, many states have adopted standardized protocols for investigating perceived clusters that are reported by citizens. Typically, this entails talking with the person who reported the cluster, verifying diagnoses

[a] Download the latest version of WinPEPI from *ww.brixtonhealth.com/pepi4windows.html*.

[b] From an economic point of view, the perception of risk has relevance. Instances in which property values have fallen following publicity about a cluster have been documented. Once a cluster is disproved, property values may return to normal (Guidotti and Jacobs, 1993).

of cases, reviewing important exposure information, and determining the extent to which the increased occurrence is ''statistically unusual.'' At each step of the investigation, findings are reported to the public and the need for more extensive and costly research is evaluated.

19.3 Goodness of fit of the Poisson distribution

Fitting the Poisson distribution

The problem of investigating a single cluster has been discussed. Thus, a more meaningful way to determine if a distribution of cases is nonrandom is to collect data over multiple years and/or locations and then compare the observed distribution of case occurrences to what is expected under a Poisson random model. When the Poisson model fits the observed distribution, the hypothesis of randomness is corroborated. When the Poisson distribution does not fit the observed frequency distribution, the hypothesis of nonrandom occurrence is supported.

Illustrative Example 19.2 Goodness of fit to Poisson distribution (''horse kicks'')

Data: Table 19.1 lists the number of fatal horse kicks in Prussian army units for the 20 years between 1875 and 1894. The unit of observation in this analysis is ''army corp-years.'' There are 200 such units of observation ($n = 200$).

Poisson model: Because the value of Poisson parameter μ is unknown, we estimate it with the sample mean ($\bar{x}$):

$$\bar{x} = \frac{\sum f_a a}{\sum f_a} \tag{19.3}$$

where f_a represents the frequency of observing a cases and $\sum f_a = n$. For the ''horse kick'' data,

$$\bar{x} = \frac{\sum f_a a}{\sum f_a} = \frac{122}{200} = 0.610$$

We thus assume $\mu = 0.610$, with Poisson probabilities calculated as:

$$\Pr(A = 0) = \frac{e^{-0.61} 0.61^0}{0!} = 0.5434$$

$$\Pr(A = 1) = \frac{e^{-0.61} 0.61^1}{1!} = 0.3314$$

and so on. Table 19.2 lists Poisson probabilities for the random model.

Expected frequencies: The next task is to determine the frequency distribution predicted by the Poisson model. The **expected frequency of observing *a* cases in a given time period** is

$$\hat{f}_a = [\Pr(A = a)][n] \tag{19.4}$$

where $\Pr(A = a)$ represents the probability of observing a cases under the Poisson model and n represents the total number of observations ($n = \sum f_a$). For example, the expected frequency of 0

fatalities in the ''horse-kick'' illustrative example is $\hat{f}_0$ = [0.5435][200] = 108.68. Table 19.3 lists the expected frequencies in column 3, and Figure 19.4 plots the observed and expected frequencies side by side. On inspection, the Poisson model fits the data well.

Goodness of fit test: The fit of the observed to expected frequencies can be tested formally. The null hypothesis H_0 is that events are randomly distributed as predicted by the Poisson model. The alternative hypothesis H_1 is that events are not randomly distributed. The null hypothesis may be false because cases are either more uniformly distributed than expected or more tightly clustered (Figure 19.5).

Before putting the data to the goodness of fit test, classes with expected frequencies of less than 1.0 are merged because the test requires that minimum expectations exceed 1 (Cochran, 1954). In the horse-kick data, we group categories of three or more fatal horse kicks to comply with this requirement (Table 19.4).

Log-likelihood goodness-of-fit statistic G: The test can be performed with a standard chi-square statistic or a *G* log-likelihood ratio statistic. The two tests yield the same conclusion when *n* is large. However, there is some advantage to the *G* log-likelihood statistic when the sample is small (Rao and Chakravarti, 1956). The log-likelihood *G* statistic is

$$G = 2\sum_{A} f_a \ln \frac{f_a}{\hat{f}_a} \tag{19.5}$$

where f_a represents the observed frequency in class *a* and $\hat{f}_a$ represents the expected frequency of class *a*. Under the null hypothesis, this statistic has a chi-square distribution with $k - 2$ degrees of freedom, where *k* represents the number of classes submitted to the test. For the data in Table 19.3, there are four classes (0, 1, 2, 3+), so $k = 4$ and df = $4 - 2 = 2$. Table 19.4 shows calculation of the *G* statistic for Illustrative Example 19.2. In this instance, $G = 0.33$ with 2 degrees of freedom. The *p* value is 0.85. Thus, we conclude that the Poisson distribution is a reasonable fit and the occurrence of fatal horse kicks over time was random.

Epidemiologic computing: WinPEPI will perform these operations with WinPEPI → Describe → C. Appraise a frequency table → 3. Values that a Poisson distribution would produce.

Table 19.1 Fatal horse kicks in the Prussian Army, 1875–1894.

Number of fatalities (*a*)	Frequency (f_a)
0	109
1	65
2	22
3	3
4	1
5+	0
Total	$\Sigma f_a = n = 200$

Source: Bortkiewicz (1898) as cited in Sokal and Rohlf (1995, p. 93).

Table 19.2 Poisson distribution, $\mu = 0.610$, fatal horse-kick illustration.

Number of fatalities	Poisson probability
0	0.5434
1	0.3314
2	0.1011
3	0.0206
4	0.0031
5	0.0004
6	0.0000
Total	1.0000

Table 19.3 Observed and expected frequencies, Illustrative Example 19.2.

Number of Fatalities *(a)*	Observed frequency (f_a)	Expected frequency ($\hat{f}_a$)
0	109	108.68
1	65	66.28
2	22	20.22
3	3	4.12
4	1	0.62
5+	0	0.08

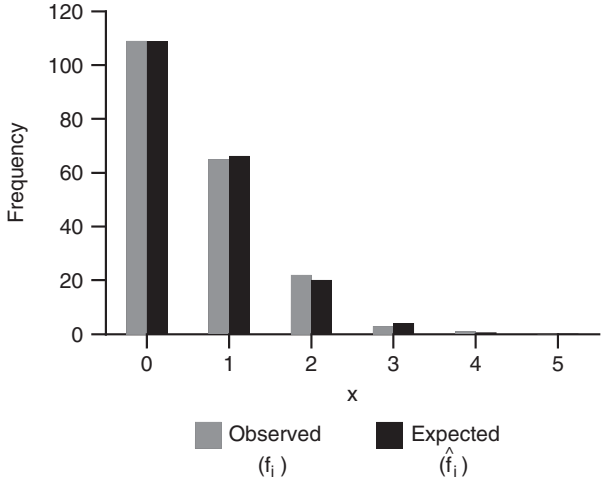


Figure 19.4 Observed and expected frequencies for Illustrative Example 19.2.

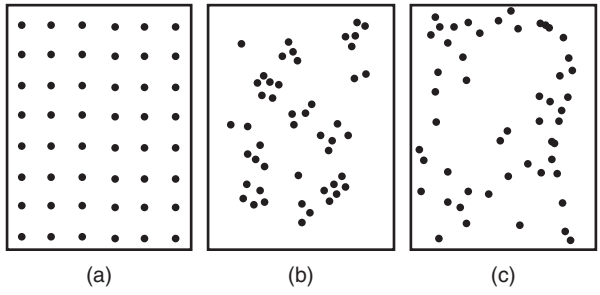


Figure 19.5 Uniform, clustered, and random distributions in space.

Table 19.4 Illustrative Example 19.2 log-likelihood ratio *G* statistic, fatal horse-kicks.

Number of fatalities *(a)*	Observed frequency (f_a)	Expected frequency ($\hat{f}_a$)	Intermediate calculation for $G^a \left(f_a \ln \frac{f_a}{\hat{f}_a} \right)$
0	109	108.68	0.320 471
1	65	66.28	−1.267 560
2	22	20.22	1.856 145
3+	4	4.82	−0.745 918
			Column sum = 0.163 138

[a] $G = 2 \sum f_a \ln \frac{f_a}{\hat{f}_a} = (2)\,(0.163\,138) = 0.326\,276.$

$\mathrm{df} = k - 2 = 4 - 2 = 2.$

$0.95 < p < 0.25$ (via Appendix Table 3).
$p = 0.85$ (via *WinPepi* → WhatIs → P value).

Illustrative Example 19.3 Frequency of war

Table 19.5 list data of the frequency of major wars in the 432 years between 1500 and 1931 (Richardson, 1944 as cited in Larson and Marx, 1981, pp. 148–149, 367). War, for this analysis, was defined as a military action that was legally declared, involved more than 50 000 troops, or resulted in a border realignment. To achieve greater uniformity in the analysis, large confrontations were split into smaller sub-wars. For example, World War I was treated as five separate wars. According to this definition, war occurred 299 times during the 432 years of observation.

We want to test whether wars are distributed randomly over time (the null hypothesis H_0) versus whether wars are *not* randomly distributed over time (the alternative hypothesis *Ha*). Calculations for the goodness-of-fit to a Poisson random model are shown in Table 19.5. There were 0.69 wars per year, on the average. The observed frequencies are shown in column 2 and the Poisson probabilities based on an expectation of 0.69 per year are shown in column 3. Column 4 shows the expected frequencies based on these Poisson probabilities. Notice that the expected frequencies are not too far off from what is observed. Intermediate calculations for the *G* log-likelihood statistics are shown in column 5. The *G* statistic is equal to 2.40 with 3 degrees of freedom for a *p*-value of 0.49). Thus, there is little evidence to reject the null hypothesis of a random distribution of wars over time. One interpretation of this analysis is that the random distribution of initiating national hostilities reflects a ''background of pugnacity'' (Richardson, 1944, p. 248).

Table 19.5 Data for Illustrative Example 19.3. Annual frequency of war in the 432 years between 1500 and 1931.[a]

Observed no. of wars a	Observed frequency	Poisson probability[b]	Expected frequency[c]	Intermediate calculation for G[d]
0	223	0.5016	216.69	6.40
1	142	0.3461	149.52	−7.33
2	48	0.1194	51.58	−3.45
3	15	0.0275	11.88	3.50
4	4	0.0055	2.38	2.08
Total	$n = \sum f_a = 432$	1.000	432	$\sum f_a \ln \frac{f_a}{\hat{f}_a} = 1.20$

Source: Richardson (1944) as cited in Larson and Marx (1981, p. 367).

[a] $\sum f_a a = (233)(0) + (142)(1) + (48)(2) + (15)(3) + (4)(4) = 299$

$\sum f_a = 432$

$$\bar{x} = \frac{\sum f_a a}{\sum f_a} = \frac{299}{432} = 0.69; \text{ use this as the estimate of } \mu$$

[b] $\Pr(A = a) = \dfrac{(e^{-0.69})(-0.69^a)}{a!}$

[c] $\hat{f}_a = \Pr(A = a)\, n$

[d] $G = 2\sum f_a \ln \dfrac{f_a}{\hat{f}_a} = (2)(1.20) = 2.40$; df $= k - 2 = 5 - 2 = 3$; pvalue $= 0.49$

Illustrative Example 19.4 Industrial accidents

The Poisson distribution will not hold when risks are not equal among the units of observation. Table 19.6 shows in the second column the frequency of accidents in women working on the manufacture of artillery shells circa 1920. The mean number of accidents was 0.465. Frequencies predicted by the Poisson function based on this mean are shown in the third column. The discrepancy between these columns suggests a poor fit. The G statistic for these data is 50.01 with 2 degrees of freedom; p-value = 1.3×10^{-11}. Thus, the population was composed of individuals with different degrees of accident proneness.

19.4 Summary

1 A cluster is a close grouping of disease or disease-related events in time or space. Many perceived clusters are, in fact, not really clusters at all, representing either ''false alarms'' or random fluctuations in occurrence. Despite the many difficulties

Table 19.6 Accidents in 647 women working on high-explosive shells in a five-week period.

Number of accidents	Observed frequencies	Expected frequencies based on the Poisson
0	447	406.31
1	132	189.03
2	42	43.97
3	21	6.82
4	3	0.79
5	2	0.08
Total	647	647

Source: Greenwood and Yule (1920, p. 275).

$\bar{x} = = \dfrac{\sum f_a a}{\sum f_a} = \dfrac{301}{647} = 0.465$ (use this as the estimate of μ)

$G = (2)(25.01) = 50.01$; df = 4 − 2 = 2; p value = 1.3×10^{-11}

encountered when investigating a cluster, most clusters are investigated for policy and legal reasons.

2 The Poisson distribution is a probability distribution that describes the random occurrence of rare discrete events in time. This probability distribution can be used to quantify chance as an explanation for clusters. Examples of Poisson calculations are presented throughout this chapter.

3 A good way to learn whether a distribution of cases is random or clustered is to use a goodness-of-fit test. Goodness-of-fit statistics compare the observed distribution of cases to a distribution predicted by a Poisson model. If the Poisson model fits the data, we conclude events are random. If the Poisson model does not fit the data, we conclude events are not randomly distributed.

Exercises

19.1 Several classes of drugs are used to treat gastric and duodenal ulcers by reducing the volume and acidity of gastric secretions. Although millions of people have been treated with these drugs and the incidence of adverse drug reactions is low, an issue of concern is that they may increase the risk of gastric cancers by causing a profoundly hypochlorhydric stomach. Suppose we find 3 cases of gastric cancer in people taking gastric acid reducers while based on the size and age distribution of the cohort, 1.2 cases were expected ($\mu = 1.2$).

(A) Calculate the probability of observing no cases in the cohort.

(B) Calculate the probability of observing one case in the cohort.

(C) Calculate the probability of observing two cases in the cohort.

(D) Calculate the probability of observing *at least* three cases in the cohort.

(E) How surprised would you be to find three or more cases under the current circumstances?

19.2 Assuming the expected number of cases in a population is 2 and events are distributed randomly, what is the likelihood of observing:

(A) no cases
(B) exactly 1 case
(C) exactly 2 cases
(D) exactly 3 cases
(E) at least 4 cases.

19.3 One of the earliest and best studied leukemia clusters in the United States occurred in Niles, Illinois, from 1956 to 1960 (Heath and Hasterlik, l963). In the 5.3-year period from 1956 to the first four months of 1961, eight cases of childhood leukemia were reported among the 7076 white children less than 15 years of age in Niles. These cases were concentrated in the St. John Brebeuf Parish. During the same period, 286 cases of childhood leukemia occurred among the 1 152 695 children in Cook County, Illinois, exclusive of Niles.

(A) Based on the rate observed in Cook County, how many cases were expected in the town of Niles?
(B) What was the probability of observing eight or more cases in Niles assuming a Poisson distribution? Do you think the eight cases in Niles represent an unusually high number of cases?

19.4 Although currently lacking in support, an ongoing theory suggests that the electromagnetic fields from cellular phones may cause brain tumors. Suppose the expected number of brain tumors in a cohort of cellular telephone users is 1.8, and 4 cases are observed. What is the probability of observing 4 or more brain tumors under random conditions?

19.5 Suppose a given stretch of highway averages 0.75 motor vehicle associated fatalities per year.

(A) What is the probability of observing 0 fatalities in a given year?
(B) What is the probability of observing 1 fatality?
(C) What is the probability of observing 2 fatalities?
(D) What is the probability of observing 3 fatalities?
(E) What is the probability of observing at least 4 motor vehicle fatalities?

References

Abramson, J.H. (2011) WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations*, **8** (1), 1 *http://www.epi-perspectives.com/content/8/1/1*.

Bender, A.P., Williams, A.N., Johnson, R.A., and Jagger, H.G. (1990) Appropriate public health responses to clusters: The art of being responsibly responsive. *American Journal of Epidemiology*, **132**, S48–S52.

Bortkiewicz, L. von. (1898) *Das Gesetz Der Kleinen Zahlen*, Tuebner, Leipzig.

Cochran, W.G. (1954) Some methods for strengthening the common χ^2 test. *Biometrics*, **10**, 417–451.

Greenberg, M., and Wartenberg, D. (1991) Communicating to an alarmed community about cancer clusters: A fifty state survey. *Journal of Community Health*, **16**, 71–82.

Greenwood, M., and Yule, G.U. (1920) An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attackes of disease or of repeated accidents. *Journal of the Royal Statistical Society*, **83**, 255–279.

Guidotti, T.L., and Jacobs, P. (1993) The implications of an epidemiologic mistake: A community's response to a perceived excess cancer risk. *American Journal of Public Health*, **83**, 233–239.

Heath, C.W., Jr., and Hasterlik, R.J. (1963) Leukemia among children in a suburban community. *American Journal of Medicine*, **34**, 796–812.

Larson, R.J., and Marx, M.L. (1981) *An Introduction of Mathematical Statistics and its Applications*, Prentice Hall, Englewood Cliffs, NJ.

Popper, K. (1959) *The Logic of Scientific Discovery* (Reprint: 2002), Routledge, London.

Rao, C.R., and Chakravarti, I.M. (1956) Some small sample tests of significance for a Poisson distribution. *Biometrics*, **12**, 264–282.

Richardson, L.F. (1944) The distribution of wars in time. *Journal of the Royal Statistical Society*, **179**, 242–250.

Rothman, K.J. (1990) A sobering start for the cluster buster's conference. *American Journal of Epidemiology*, 132 (Suppl. 1), S6–Sl3.

Sokal, R.R., and Rohlf, F.J. (1995) *Biometry. The Principles and Practice of Statistics in Biologic Research*, 3rd edn, W.H. Freeman, New York.

Answers to Exercises and Review Questions

Chapter 1: Epidemiology past and present

Review questions

R.1.1 *epi* = ''upon''; *demos* = ''the people''; *ology* = ''to speak of'' or ''to study.''

R.1.2 A personal response is requested.

R.1.3 There are several differences between epidemiology and clinical medicine. One difference is their primary unit of concern. The primary unit of concern in epidemiology is the group (an ''aggregate of individuals''). In clinical medicine, the primary unit of concern is the individual. With respect to epidemiology and public health: epidemiology is primarily a ''study of,'' while public health is an ''activity'' requiring social participation.

R.1.4 Elements of the 1948 WHO definition of health and well-being: (1) physical, (2) mental, (3) social.

R.1.5 See Table 1.1.

R.1.6 The epidemiologist must communicate their findings in order to effectively participate with other disciplines and sectors in deciding and implementing public health practices and interventions.

R.1.7 Morris's seven uses of epidemiology: (1) historical study; (2) community diagnosis; (3) workings of health services; (4) individual chances; (5) complete the clinical picture; (6) identify syndromes; (7) search for causes.

R.1.8 Community diagnosis determines the incidence and prevalence of disease and disease determinants in communities and community subgroups.

R.1.9 Key elements of demographic transition of the 20th century: increased longevity, decreased fertility, aging of the population.

R.1.10 Key elements of the epidemiologic transition of the 20th century: decreases in acute and contagious diseases; increases in chronic, noninfectious, lifestyle diseases.

R.1.11 Steep declines have been seen in cardiovascular disease (heart attacks), cerebrovascular disease (strokes), and pneumonia and influenza.

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

R.1.12 The population pyramid has ''squared off'' (i.e., became less of a pyramid) over time, with a larger percentage of the population shifted toward the older age groups.

R.1.13 Examples of modifiable risk factors: tobacco use, alcohol use, diet, high blood pressure, high risk sexual practices, exposure to sunlight and other forms of radiation.

R.1.14 False. While it is true that death due to some cancers increased during the 20th century (e.g., lung cancer), others have declined. In addition, much of apparent increases have been due to the aging of the population. Thus, after age-adjustment, the ups and downs more-or-less balanced out resulting in a fairly flat cancer mortality rate (see Figure 1.2).

R.1.15 False. Age-adjusted cardiovascular mortality rates continue to decline.

R.1.16 Heart disease; cancer; stroke.

R.1.17 True.

R.1.18 Epidemiology became a recognized discipline in the 19th century with the creation of the Epidemiological Society of London (established 1850).

R.1.19 Hippocrates (400 BCE).

R.1.20 Measuring, sequencing, classifying, grouping, confirming, observing, formulating, questioning, identifying, generalizing, experimenting, and testing.

R.1.21 Matching: A = Syndenham; B = Pott; C = Graunt; D = Fracastoro; E = Salmon; F = Pinel; G = Louis; H = Farr; I = Snow.

Chapter 2: Causal concepts

Exercises

2.1 No answer provided.

2.2 d

2.3 c

2.4 b

2.5 a

2.6 c

2.7 b

2.8 c

2.9 b

2.10 a

2.11 Matching:

a = Virulence
b = Sufficient constellation
c = Non-necessary Component Cause

d = Infectivity
e = Causal web
f = Pathogenicity
g = Necessary cause

2.12 Matching:
a = Coherence
b = Biologic gradient
c = Plausibility
d = Temporality
e = Experimentation
f = Analogy
g = Specificity
h = Consistency
i = Strength

2.13 Matching:
a = Biologic gradient
b = Plausibility
c = Strength
d = Consistency
e = Analogy
e = Temporality

2.14 Consistency implies that studies are consistent in the estimation of association. If an exposure consistently causes disease in a given person, it is said to be sufficient. Sufficiency is not one of Hill's criteria for causality.

Review questions

R.2.1 Stages of disease: susceptibility, preclinical, clinical, resolution (recovery, disability, or death).

R.2.2 The onset of the preclinical stage is exposure to the agent. The onset of the clinical stage is marked by first symptoms. The onset of the resolution is recovery, disability or death.

R.2.3 The objective of primary prevention is to prevent new occurrences. The objective of secondary prevention is to delay the onset of disease or decrease its severity. The objective of tertiary prevention is to slow progression or minimize the progression of disease.

R.2.4 The agent multiplies within the host during the incubation of an infectious disease.

R.2.5 Synonyms for *incubation period*: latent period, empirical induction period (roughly).

R.2.6 Tuberculosis, AIDS, leprosy.

R.2.7 Mammography is a form of secondary prevention; because it detects disease after it has been initiated but before it becomes clinical.

R.2.8 There are many reasons it is important to understand the natural history of HIV for its effective control. As an example, we must be aware of the period between

infection and detection of antibodies to understand that diagnosis of the disease will be delayed until some time after exposure. In addition, one must be aware of the long incubation period during which the host is symptom free but is still contagious.

R.2.9 The spectrum of disease. (For infectious diseases: the gradient of infection.)

R.2.10 This describes the epidemiologic iceberg.

R.2.11 A cause is any predecessor without which the effect would not have occurred or would have occurred at a later time.

R.2.12 A causal interaction is the bringing about of an effect by two or more factors act together.

R.2.13 Measles virus is necessary but not sufficient to cause measles. (The agent will not cause disease in an immune individual.)

R.2.14 A causal mechanism is completed when the outcome becomes inevitable.

R.2.15 The causal complement is (E + F).

R.2.16 The causal complement is E.

R.2.17 Phenylketonuria is both a genetic disease and an environmental disease. The genetic disorder involves the deficiency in the enzyme needed to metabolize phenylalanine. The environmental component is the presence of phenylalanine in the diet.

R.2.18 Contributing/component causes of hip fractures in the elderly: Low calcium diet, osteoporosis, genetic susceptibility to osteoporosis, female sex, weakness, poor balance, sedation, slippery surface, lack of hand rails, etc.

R.2.19 A direct cause is close to the pathogenic mechanism. An indirect cause is connected to the pathogenic through other factors.

R.2.20 Types of pathogenic agents: biological, physical, and chemical.

R.2.21 Types of chemical pathogenic agents: nutritive excesses and deficiencies, toxins, drugs, allergens.

R.2.22 Types of physical pathogenic agents: heat, light, radiation, noise, vibration, and objects that cause trauma.

R.2.23 Infectivity = ability to infect; pathogenicity = ability to cause disease; virulence = ability to cause severe disease.

R.2.24 Epidemiologic homeostasis occurs when agent, host, and environmental causes of disease are balanced in such a way as to maintain the current rate of disease in the population.

R.2.25 Causal inference is the process of deriving cause-and-effect conclusions from fact and knowledge.

R.2.26 We base preventive measures on knowledge of causal mechanisms to increase their efficacy. False knowledge can have a contrary effect.

R.2.27 There are times when discovery of effective preventive measures pre-date identification of the causal mechanism

R.2.28 See Table 2.3.

R.2.29 This initial Surgeon General's Report on Smoking and Health was published in 1964.

R.2.30 Because there are always alternative explanations for associations.

R.2.31 False. Statistical methods cannot by themselves establish proof.

R.2.32 Ratios of incidences (relative risks) are generally considered to be the most direct measure of the strength of an epidemiologic relationship.

R.2.33 Strong associations are less likely to be ''explained away'' by confounding.

R.2.34 No. Weak associations are just more difficult to ''prove.''

R.2.35 No, since multiple studies may be consistently incorrect, especially if they share or exhibit multiple flaws.

R.2.36 True. This is the *sine quo non*. The cause must always precede the effect.

R.2.37 *Coherence* holds that all sources of evidence ''stick together.'' *Plausibility* holds that relations can be explained by current knowledge.

R.2.38 This is an example of an analogy.

R.2.39 Biological gradient.

R.2.40 Consistency, strength, specificity, temporality, biological gradient, plausibility, coherence, experimentation, analogy.

Philosophical considerations

R.2.41 Ultimate proof in empirical sciences is not possible. However, statements of proof can be very strong, and even overwhelming.

R.2.42 Decisions having to do with scientific hypotheses (type 1) require rigorous skepticism. The latter having to do with public health interventions (type 2) may require making a reasonable choice based on available information.

R.2.43 The Problem of Induction is the philosophical quandary that observed sequences of occurrence do prove cause and effect (*post hoc propter hoc*).

R.2.44 True. Refutationists believe that a theory is not scientific unless it is falsifiable.

R.2.45 One can never fully prove that all swans are white because the next swan that comes long may be black or even light gray.

R.2.46 No amount of observations can prove a hypothesis. In contrast, one strong disproof can dispel a theory.

Chapter 3: Epidemiologic measures

Exercises

3.1 Point prevalences, period prevalence, and risk

(A) 4/10 = 0.40

(B) 2/10 = 0.20

(C) 6/10 = 0.60

(D) 2/5 = 0.40 Two disease onsets (A and E) observed among five individuals at risk (A, D, E, F, and I).

3.2 Hypertension in a cohort of men

(A) $\dfrac{50 \text{ people}}{1000 \text{ people}} = 0.05$

(B) $\dfrac{64 \text{ people}}{(1000-50) \text{ people}} = 0.0674$

(C) With the actuarial adjustment: $\dfrac{64}{(64 \times 2.5) + (886 \times 5)} = 0.0139 \text{ year}^{-1}$

Without the actuarial adjustment: $\dfrac{64}{(950 \times 5)} = 0.0135 \text{ year}^{-1}$

The adjustment made little difference because the outcome was uncommon.

3.3 Vital statistics

All rates are ''per 1000.''

(A) 30

(B) 15

(C) 17

(D) 40

(E) 3

(F) 1

3.4 Prevalence in an open population

(A) Increase

(B) Decrease

(C) Increase

(D) Decrease

(E) Decrease

3.5 More vital statistics

(A) Birth rate per 1000 person-years = $\dfrac{4\,065\,014}{255\,078\,000} \times 1000 = 15.9$

(B) Mortality rate per 100 000 population = $\dfrac{2\,175\,631}{255\,078\,000} \times 100\,000 = 853$

(C) Infant mortality rate per 1000 live births = $\dfrac{34\,553}{4\,065\,014} \times 1000 = 8.50$

3.6 Effect of a Treatment An effective treatment that increases survival but does not result in a permanent cure would have no influence on incidence, but would increase prevalence over time.

3.7 Fatalities associated with travel

(A) Traffic fatality rate, 1993 = $\dfrac{40\,100}{2297 \times 10^9} = 1.746 \times 10^{-8}$ miles travels

(B) The fatality rate per miles traveled during scheduled air transportation (5 per billion miles traveled) is (1/35)th the per mile rate associated with motor vehicle travel (17.46 per billion miles traveled).

3.8 Accidents in hospitals The author's interpretation is incorrect. The data represent incidence counts; no ''denominator data'' are presented. Such data cannot be used for statements about rates or risks. It is possible that there are many more patients in the 62 and over age group than in any other age group,

and the higher number of accidents simply reflects this great number of people initially at risk.

3.9 Stationary? The size of the US population is increasing and its age distribution is shifting. Therefore, US demographics are not stationary.

3.10 Mortality rate and life expectancy

In population 1: R = 2/150 years = 0.013 33 year^{-1} and life expectancy = (1/0.013 33 year^{-1}) = 75 years, which is also the average age at death in this closed fully-followed cohort.

In population 2: R = 2/175 years = 0.011 43 year^{-1} and life expectancy = (1/0.011 43 year^{-1}) = 87.5 years = average age at death.

3.11 Comparing prevalences We can not conclude population A has twice the risk because prevalence depends on both incidence and mean duration of disease. If the cases in population A survived twice as long as the cases in population B, it could have the same incidence and double the prevalence.

3.12 Risk and rate of Breast cancer

(A) Risk (incidence proportion) = 250/9500 = 0.026 32 = 26.3 per 1000

(B) Person-years at risk $\approx$ (9500 $\times$ 5) years $-$ (250 $\times$ 2.5) years = (47 500 $-$ 625) years = 46 875 person-years. Therefore, the Rate = 250/46 875 years = 0.005 33 year^{-1} = 5.33 per 1000 years

(C) Risk $\approx$ rate $\times$ time = (5.33 per 1000 years) $\times$ 5 years = 26.65 per 1000

3.13 Coronary heart disease

(A) Risk = 100/800 = 0.125 = 12.5 per 100

(B) Rate = 100/[(800 $-$ 100) $\cdot$ 10 + (100 $\cdot$ 5)] =0.0133 per year = 1.33 per 100 person-years

3.14 Driving errors

(A) $\dfrac{1 \text{ mistake}}{2 \text{ miles}} = {}^1\!/_2$ mistake per mile

(B) $\dfrac{1 \text{ (mistaken) observation}}{400 \text{ observations}} = 0.0025$

(C) $\dfrac{1 \text{ mistake}}{500 \text{ miles}} \times \dfrac{61\,000 \text{ miles}}{1 \text{ crash}} = \dfrac{122 \text{ mistakes}}{\text{crash}}$

3.15 N = 6

(A) Sum of person-time = 4 + 4 + 2 + 3 + 9 + 6 = 28

(B) Average number of people = 28 person-years/10 years = 2.8 people

(C) Rate $= \dfrac{\text{onsets}}{\text{average pop'n size} \times \text{time}} = \dfrac{4}{2.8 \text{ people} \times 10 \text{ years}} = 0.143$ per person-year

3.16 Cohort study

(A) Prevalence = 10/150 = 0.0667

(B) Prevalence = (10 + 16)/150 = 0.1733

(C) Incidence proportion = 16/(150 − 10) = 0.1143
(D) Incidence rate = 16/[(140 − 16) · 5 + (16 · 2.5)] = 0.0242 per year or 24.2 per 1000 years

3.17 More population based rates
(A) 12
(B) 10
(C) 10
(D) 30

3.18 Actuarial adjustment of person-Time (18 persons × 1 year) + (82 persons × 1 years) = 182 person-years

3.19 Framingham men $RR = R_1/R_0 = 0.1203/0.0352 = 3.4$. Here is how the original paper (Kannel *et al.*, 1961, p. 39) reported the results: ''Analysis of these groups reveals a gradient of risk of developing CHD, such that those with serum cholesterol over 244 mg per 100 ml have more than three times the incidence of CHD as those with cholesterol levels less than 210 mg per 100 ml.''

3.20 Framingham women $RR = R_1/R_0 = 0.0435/0.0180 = 2.4$. The high cholesterol women had more than double the risk of CHD.

3.21 Restenosis
(A) *RR* = (21/49)/(2/26) = 5.6
(B) 2-by-2 table

	Restenosis +	Restenosis −	
Cytomegalovirus +	21	28	49
Cytomegalovirus −	2	24	26

(C) *OR* = (21/28)/(2/24) = 9.0. The *OR* exceeded the *RR* because the outcome (restenosis) was no uncommon.

3.22 Primary cardiac arrest and vigorous exercise
RR associated with 1–19 min/week = 14/18 = 0.78
RR associated with 2–139 min/week = 6/18 = 0.33
RR associated with ≥140 min/week = 5/18 = 0.28
The *RR*s show progressive declines with increasing levels of habitual high-intensity activity.

3.23 California mortality
(A) cR (per 100 000) = 215 111/30 381 000 × 100 000 = 708 per 100 000
This rate is substantially lower than Florida's crude death rate of 1026 per 100 000.

(B) Calculation of the age-adjusted death rate for California, 1991

Stratum (i)	Age (years)	Deaths	Population	Rate (per 100 000) (r_i)	Standard million (N_i)	Product ($N_i r_i$)
1	0–4	5500	2 651 000	207	76 158	15 800 415
2	5–24	5736	8 824 000	65	286 501	18 623 864
3	25–44	19 178	10 539 000	182	325 971	59 317 505
4	45–64	37 313	5 179 000	720	185 402	133576073
5	65–74	45 306	1 874 000	2418	72 494	175 262 175
6	75+	102 078	1 314 000	7768	53 474	415 412 403
Column sums →		215 111	30 381 000	708	1 000 000	817 992 435

Calculations:

$$\sum N_i r_i = N_1 r_1 + N_2 r_2 + \cdots + N_6 r_6 = 15\,800\,415 + 18\,623\,864 + \cdots + 415\,412\,403 = 817\,992\,435$$

$$\sum N_i = N_1 + N_2 + \cdots + N_6 = 76\,158 + 286\,501 + \cdots + 53\,474 = 1\,000\,000$$

$$\text{aR}_{\text{direct}} = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i} = \frac{817\,992\,435}{1\,000\,000} \approx 818$$

(C) $\text{aR}_{\text{direct}}$ = 818 per 100 000.

(D) California's age-adjusted death rate is slightly higher than Florida's adjusted death rate of 784 per 100 000.

3.24 Arkansas mortality

(A) cR (per 100 000) = 25 045/2 372 000 × 100 000 = 1056 per 100 000. This rate is similar to Florida's crude death rate of 1026 per 100 000.

(B) Calculation of the age-adjusted death rate for Arkansas, 1991

Stratum (i)	Age (years)	Deaths	Population	Rate (per 100 000) (r_i)	Standard million (N_i)	Product ($N_i r_i$)
1	0–4	449	170 000	264	76 158	20 114 672
2	5–24	562	697 000	81	286 501	23 100 941
3	25–44	1459	694 000	210	325 971	68 529 062
4	45–64	4,072	458 000	889	185 402	164 837 761
5	65–74	5466	196 000	2789	72 494	202 169 492
6	75+	13 037	157 000	8304	53 474	444 038 559
Column sums →		25 045	2 372 000	1056	1 000 000	922 790 487

Calculations:

$$\sum N_i r_i = N_1 r_1 + N_2 r_2 + \cdots + N_6 r_6 = 20\,114\,972 + 23\,100\,941 + \cdots + 444\,038\,559 = 922\,790\,487$$

$$\sum N_i = N_1 + N_2 + \cdots + N_6 = 76\,158 + 286\,501 + \cdots + 53\,474 = 1\,000\,000$$

$$\text{aR}_{\text{direct}} = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i} = \frac{922\,790\,487}{1\,000\,000} \approx 923$$

(C) aR_{direct} = 923 per 100 000.

(D) Arkansas's age-adjusted death rate is substantially higher than Florida's adjusted death rate of 784 per 100 000.

3.25 Egyptian mortality

(A) cR = 416 000/55 163 000 = 0.007 54 = 754 per 100 000. This rate is substantially lower than that of the United States.

(B) $aR_{indirect}$ = 1849 per 100 000 This mortality rate is more than twice that of the United States (SMR = 2.15).

Calculation of adjusted rate by the indirect method for Egypt, using the United States, 1991, as the external reference population

Age (years)	US rate (R_i)	Egyptian population (n_i)	Product ($R_i n_i$)
0–4	0.00229	7 909 000	18 112
5–24	0.00062	24 560 000	15 227
25–44	0.00180	13 764 000	24 775
45–64	0.00789	6 921 000	54 607
65–74	0.02618	1 485 000	38 877
75+	0.08046	524 000	42 161
Total		55 163 000	193 759

cR = 416 000/55 163 000 = 0.007 54 = 754 per 100 000.

$$\mu = \sum R_i n_i = R_1 n_1 + R_2 n_2 + \cdots + R_6 n_6$$
$$= 18\,112 + 15\,227 + \cdots + 42\,161 = 193\,759$$
$$\text{SMR} = \frac{x}{\mu} = \frac{416\,000 \text{ (see Table 7.14)}}{193\,759} = 2.15$$
$$aR_{indirect} = (\text{cR})(\text{SMR}) = (0.008\,60)(2.15)$$
$$= 0.018\,49 = 1849 \text{ per } 100\,000$$

Review questions

R.3.1 The numerator of incidences includes only cases that had onsets during the period of observation. Prevalence counts all cases, old and new.

R.3.2 Not necessarily. New York has a larger population. The greater number of deaths may merely reflect its large population size.

R.3.3 Size, time of observation, age, and other and characteristics

R.3.4 A ratio is a combination of two numbers that show their relative sizes. It is one number divided by another.

R.3.5 Cohort.

R.3.6 Average risk, risk, cumulative incidence.

R.3.7 Numerator = no. of disease onsets; denominator = size of cohort at risk.

R.3.8 Because the objective of an incidence proportion is to estimate the probability of developing the disease.

R.3.9 (a) The period length of observation. (b) The age distribution of the group.

R.3.10 40.

R.3.11 Incidence density, average hazard, person-time rate.

R.3.12 1 person observed for a year; 2 people observed for half a year each; 3 people observed for one-third of a year each; *etc*.

R.3.13 68 person-hours.

R.3.14 Numerator = no. of disease onsets; denominator = amount of ''person-time'' in population.

R.3.15 0.013 33 person-year^{-1} $\times$ 1000 person-years = 13.3.

R.3.16 When the disease is rare (cumulative incidence $<5\%$) and the period of observation is one year.

R.3.17 Numerator = number of existent cases; denominator = population size.

R.3.18 It considers both new and old cases and involves no follow-up of individuals.

R.3.19 It will increase.

R.3.20 ''Exposure'' (independent variable) and ''disease'' (dependent variable).

R.3.21 True.

R.3.22 It means that higher levels of exposure are associated with *higher* incidences of disease.

R.3.23 Division.

R.3.24 Subtraction.

R.3.25 This is an example of an *RD*.

R.3.26 This is an *RR*.

R.3.27 *RD* quantifies the effect of the exposure in absolute terms.

R.3.28 *RR* quantifies the effect of the exposure in relative terms.

R.3.29 *RR*.

R.3.30 It would not be correct to make this statement because the exposure increases risk by only 50%.

R.3.31 False.

R.3.32 True.

R.3.33 It changes to it's reciprocal. For example, an *RR* of 2 becomes 1/2 = 0.5.

R.3.34 85%.

R.3.35 AF_e.

R.3.36 AF_p.

Chapter 4: Descriptive epidemiology

Exercises

4.1 Ecological correlations There are strong positive correlations for cigarettes and bladder cancer and cigarettes and lung cancer; there is a moderately strong correlation for cigarettes and kidney cancer. There are associations between bladder cancer, lung cancer, and kidney cancer. The implication is that lung cancer, kidney cancer, and bladder cancer may have a common underlying cause, perhaps cigarettes or something associated with cigarettes.

4.2 Notifiable conditions No answer provided.

4.3 ***Le Suicide***

(A) Here are four observations from Durkheim: (1) Marriage before the age of 20 (''too early marriages'') has an aggravating influence on suicide, especially in men. (2) After age 20, married persons of both sexes enjoy some protection from suicide in comparison with unmarried people. (3) The protective effect of marriage is greater in men than in women. (4) Widowhood diminishes the protective effects of marriage but does not entirely eliminate it.

(B) Durkheim reflected on whether the apparent protective effects of marriage were due to the influence of the married domestic environment or whether this ''immunity'' was due to some sort of ''matrimonial selection'' in which people who marry have certain physical and moral constitutions that make them less likely to commit suicide. This type of reflective reasoning and careful interpretation foreshadows the modern epidemiologic approach.

Review questions

R.4.1 Descriptive epidemiology explores rates according to person, place, and time variables with the primary intention of generating hypotheses. Analytic epidemiology collects data that has been specifically designed to address hypotheses about specific risk factor.

R.4.2 False. There is no firm demarcation between descriptive epidemiology and analytic epidemiology.

R.4.3 A case series is a description of the history and clinical manifestations of a small number of individuals with a particular disease outcome.

R.4.4 Case series lack the denominator data needed to calculate rates.

R.4.5 True, for example see Illustrative Example 4.1.

R.4.6 Epidemiologic surveillance systems are structures set up to collect and analyze outcome-specific health data for planning, carrying out, and evaluating public health practices.

R.4.7 Active surveillance requires an active seeking out of population-based cases. Passive surveillance relies on doctors, hospitals, and the public to send reports to the appropriate public health surveillance system voluntarily.

R.4.8 The National Center for Health Statistics

R.4.9 place, and time.

R.4.10 Person.

R.4.11 Host factor closely tied to place: cultural practices, occupation, recreational practices, etc.

R.4.12 Environmental factor that is closely tied to ''place'': climate, economic development, etc.

R.4.13 Studies show that Japanese–American women develop breast cancer rates that are typical of American women after several generations of acculturation.

R.4.14 (a) Propagating epidemic, (b) point epidemic, (c) endemic, (d) sporadic.

R.4.15 A unit of observation is the level of human aggregation upon which measurements are recorded.

R.4.16 ecological

R.4.17 True.

R.4.18 An ecological correlation is a correlation in which the units of observation are based on group rather than individual characteristics.

R.4.19 Neighborhood crime rate is an integral aggregate-level variable.

R.4.20 True.

R.4.21 An ecological fallacy (also called aggregation bias) occurs when an association seen in aggregate data does not apply to individuals.

R.4.22 A multilevel study incorporates individual- and aggregate-level variables in order to help untangle relationships between direct and indirect causes of disease.

R.4.23 Confounding bias is a spurious association caused by extraneous factors.

R.4.24 (1) Contextual variable, (2) integral variable, (3) contagion variable.

R.4.25 This is an integral group property.

Chapter 5: Introduction to epidemiologic study design

Exercises

5.1 Study types.

(A) The exposure is working with poultry. The study outcome is being positive for the avian A-V antibody. This is a cross-sectional observational study because it is observational based on exposure groups, but we cannot accurately discern date of seroconversion of cases.

(B) The exposure is type A or B behavior. The study outcome is coronary reoccurrence. This study is an observational cohort.

(C) The exposure is bus driver vs. office worker. The study outcome is hypertension. This study is cross-section because the design does not permit us

to discern whether hypertension began before or after the subjects entered their current job.

(D) The exposure is high dietary fat consumption. The study outcome is breast cancer. This is a case–control study because subjects were selected based on disease status.

(E) The exposures are various personal characteristics (e.g., job category, blood pressure, diet type, exercise program). The study outcome is the onset of coronary heart disease. This is an observational cohort study.

(F) The exposure is coaching to pursue regular moderate exercise. The study outcome is unspecified cardiac health measures. This is an experimental study.

(G) The study exposure is eating raw clams and oysters over the preceding year. The study outcome is infectious hepatitis. This is a case–control study.

(H) The study exposure is smoking. The study outcome is musculoskeletal symptoms. This is a cross-sectional study.

(I) The study exposure is size of the manufacturing plant. The study outcome is accident rates. It is unclear from the description whether the unit of observation is the individual or the manufacturing plant. Thus, the study is either an ecological study or observational cohort study.

5.2 Driving while talking.

(A) The exposure is the amount of talking on the phone while driving. This exposure can be measured in, say, minutes per month, or can be classified as say, ''non-talkers,'' ''light talkers,'' and ''heavy talkers.'' Data could possibly come from self-reporting, from reporting by spouses or roommates, or from cell phone records.

(B) The cases will be drivers that died in automobile accidents. This information could be derived from coroner's reports, police records, or death certificates.

(C) We would strive to record information on anything that might increase the risk of traffic fatalities, such as amount driven, road conditions, drivers' age, sex, and ethnicity, use of seat belts, medical history, and other factors thought to be associated with cell phone use influence while driving and the risk of fatal automobile accidents.

(D) Information on phone usage and other variables may be difficult to obtain accurately, especially in cases.

(E) Cases would be automobile fatalities. Controls would be living drivers selected from the same source population as the cases, preferably via random sampling all drivers in the region. Amount talked on the phone while driving could be based on historical phone records. Information on the extraneous variables identified in part C would also be obtained retrospectively.

5.3 Agricultural injuries

(A) The exposure variables are race (classified as Caucasian or African–American) and farm ownership (classified as owner or worker).

(B) The outcome variable is agricultural related injury.

(C) The exposure variables being studied (race, ownership status) are not ''assignable.''

5.4 Open population rates Studies in open populations are unable to following individual experience over extended periods of time and, therefore, are not longitudinal.

5.5 Classify the study

(A) Exposure = chloramphenicol, Disease = Rocky Mountain spotted fever, Design = case report. Case reports are not always regarded as analytic studies. Nevertheless, an astute case report can serve to alert us of an important public health concern.

(B) Exposure = previous radiation exposure, Disease = thyroid cancer, Design = case series (no referent group).

(C) Exposure = chewing tobacco, Disease = carcinoma of the stomach, Design = Case–control study.

(D) Exposure = median income, Disease = pollution level, Design = ecological (unit of observation is the census tract).

(E) Exposure = oral contraceptives, Disease = fatal circulatory system disease, Design = observational cohort.

(F) Exposure = school breakfast program, Study outcome = height and weight, Design = experimental (community trial).

Review questions

R.5.1 Experimental studies assign study exposures to subjects according to rules set by the study protocol (e.g., random assignment of exposure). Observational studies classify exposures in individuals as they already exist.

R.5.2 Randomization balances groups with respect to would-be confounding factors. (Randomization increases the likelihood that differences in study outcomes at the end of the study can be attributed to the treatments and not to confounding factors by achieving equal distributions of known and unknown confounding factors in the exposed group and nonexposed group.)

R.5.3 ''Current health status'' is cross-sectional.

R.5.4 Longitudinal studies monitor individual experience over time. Cross-sectional studies address subjects at a single point in time or over a brief period of time and do not permit the accurate placement of events on a timeline within individuals.

R.5.5 Cohort studies start with disease-free individuals and tracks health incidents over time. Case–control studies start with diseased and non-diseased study subjects and then ascertain prior exposure to risk factors.

R.5.6 Ecological studies use aggregate-level data (only).

R.5.7 (1) = Ecological, (2) = Cross-sectional study, (3) = Case–control, (4) = Cohort, (5) = Experimental.

R.5.8 False. Longitudinalness depends on whether your events in individuals can be accurately placed in time. The time-element described in the question is called ''proximity.''

R.5.9 (1) = prevalence-incidence bias, (2) = detection bias, (3) = reverse-causality bias (cart-before-the-horse bias).

R.5.10 Equipoise is balanced doubt.

R.5.11 Informed consent is the ability to give consent to participate or not participate in a study freely in light of the facts and without obligation.

R.5.12 Respect for individuals, beneficence, and justice.

R.5.13 IRB stands for Institutional Review Board.

R.5.14 IRBs ensure the ethical conduct of studies.

Chapter 6: Experimental studies

Exercises

6.1 Bicycle helmet campaign

(A) The unit of intervention is the elementary school.

(B) The unit of observation is the individual bicycle rider.

(C) We could block (stratify) schools based on socioeconomic status and then randomize the intervention within blocks.

6.2 Five cities

(A) This study is experimental because it included an intervention imposed by the study's protocol.

(B) This study is a community trial because the interventions were made environmentally, and were not delivered on an individual-by-individual basis.

(C) Ascertainments were made blindly to prevent differential misclassification from biasing the results.

(D) Rates of the outcome were declining over time in both treatment and control cities. Had there been no control cities, it might have appeared as if the declines in the treatment cities were due to the interventions.

6.3 Fictitious HIV vaccine trial

(A) This study is a primary prevention field trial because a preventive intervention was applied to individuals.

(B) General ethical concerns include respect for individuals, beneficence, justice informed, consent, IRB oversight, equipoise, and oversight by an data and safety management board. Specific concerns include awareness of the potential for adverse reactions to the vaccine and the possibility of creating a sense false sense of security in participants that could encourage unwarranted risk taking behavior.

(C) Per-protocol $RR = R_1/R_0 = (4/154)/(11/163) = 0.02597/0.06748 = 0.38$. Efficacy $= 1 - 0.38 = 0.62$.

(D) Intention-to-treat $RR = (4/200)/(11/200) = 0.0200/0.0550 = 0.36$. Efficacy $= 1 - 0.36 = 0.64$.

(E) This does not materially change our conclusion: either way the vaccine is more than 60% effective.

Review questions

R.6.1 *Trier* means ''to try'' in French.

R.6.2 An experiment assigns the study exposure to subjects; often this assignment will be based on chance mechanisms (a randomized trial). Observational studies study the effects of the study exposure without intervention in a ''come as you are'' fashion.

R.6.3 In clinical trials, therapeutic interventions are administered to individuals. In field trials, preventive interventions are administered to individuals. In community trials, therapeutic or preventive interventions are administered at the aggregate level (e.g., a health education program directed to groups).

R.6.4 Randomized ≡ the study treatment assigned via chance mechanisms.
Controlled ≡ the treatment group is compared to a control group.
Double-blinded ≡ study subjects and outcome ascertainers are kept in dark about which treatment is received by each subject.

R.6.5 Lind's allocation of treatments were not randomized.

R.6.6 Parés had run out of the standard treatment and was forced to switch to an alternative treatment. Therefore, assignment of the intervention was an ''act of nature'' and was not under the control of the study protocol.

R.6.7 To provide a meaningful baseline for comparison.

R.6.8 This is The Hawthorne Effect.

R.6.9 The placebo effect is perceived improvements in a condition associated with the use of an inert intervention.

R.6.10 Randomization reduces confounding by balancing the treatment and control group with respect to measured and unmeasured extraneous factors that are associated with the response. This encourages ''like-to-like'' comparisons and reduces the opportunity for confounding.

R.6.11 Admissibility criteria are the conditions that determine whether an individual is eligible to participate in a study.

R.6.12 Intention-to-treat analysis considers outcomes in participants that withdraw and whether noncompliant.

R.6.13 Intention-to-treat analysis offers these advantages: (a) it more accurately reflects the way the treatment will perform under real conditions and (b) it simplifies the task of guarding against conscious and unconscious attempts to influence the results of the study.

R.6.14 A synonym for *intention-to-treat analysis* is *analyze-as-randomized*.

R.6.15 Evidence of group comparability is necessary even in randomized experiments because randomization does not guarantee comparability, especially when the study is small.

R.6.16 Reasons randomization does not guarantee group comparability: (a) random difference (especially when the study is small), (b) flaws in the randomization procedure, (c) differences inadvertently imparted by the intervention.

R.6.17 Even though blinding does not guarantee accurate ascertainment of study outcomes, it is still useful because errors will be random and will tend to occur equally in the treatment and control group. This will reduce the potential for information bias.

Chapter 7: Observational cohort studies

Exercises

7.1 Farm work injuries

(A) This is a prospective observational cohort study. Subjects were contacted biannually to ascertain outcomes.

(B) Rate ratios are reported in the last column of this table:

Group	Cases	Person-years	Rate per 1000 person-years	Rate ratio
Caucasian owners	67	2047	32.7	**Referent**
African–American owners	27	821	32.9	**1.01**
African–American workers	37	359	103.1	**3.15**

(C) Race does not appear to be an independent risk factor, since the African–American farm owners and Caucasian farm owners experienced similar rates. However, lack of farm ownership does appear to be an independent risk factor because African–American workers experienced accidents at about three times the rate of African–American farm owners.

7.2 NSAIDS and breast cancer

(A) Ethics demands that human subjects freely provide informed consent before being participating in a study.

(B) The exposure to pain medications is self-reported and somewhat technical. It therefore needs to be validated.

(C) This study is ambidirectional because exposure information was collected retrospectively and breast cancer outcomes were ascertained prospectively.

(D) To prevent the known exposure status of individuals from biasing the coders when classifying disease status.

(E) $RR_1 = \dfrac{R_1}{R_0} = \dfrac{4.55}{4.90} = 0.93$

(F) $RR_2 = \dfrac{R_2}{R_0} = \dfrac{4.06}{4.90} = 0.83$

(G) Aspirin use appears to decrease the risk breast cancer in a dose–response fashion.

7.3 Oral contraceptive estrogen dose (PYs = person-years)

(A) $R_1 = \dfrac{69}{9.8 \times 10^4 \text{ PYs}} \times 10\,000 = 7.04$ per 10 000 PYs

$R_0 = \dfrac{53}{12.7 \times 10^4 \text{ PYs}} \times 10\,000 = 4.17$ per 10 000 PYs

(B) RD = (7.04 per 10 000 PYs) − (4.17 per 10 000 PYs) = 2.87 per 10 000 PYs

(C) A little less than 3 (2.87 to be exact).

(D) To prevent them from differentially misclassifying cases based on preconceived notion about the risks associated with different formulations in order to prevent information bias.

7.4 Angry-prone personality and coronary heart disease

(A) The exposure is anger-prone personality as measured by the Spielberger anger-temperament trait scale. The disease was coronary disease incidents.

(B) An anger-prone temperament is regarded as a semi-immutable personal characteristic. An effective clinical trial would have to induce a person to either become more anger-prone (which would probably be unethical) or less anger-prone (which would be a major challenge).

(C) Cohort studies require investigators to measure the incidence of events. Prevalent cases would therefore need to be eliminated before follow-up began.

(D) Although losing only 7% of the cohort is considered to be an acceptably low loss to follow-up rate, we must still be aware of the potential for selection bias if the loss to follow-up was associated with both the exposure and disease.

(E) These factors are associated with the exposure and disease and may therefore confound observed associations derived from this study.

(F) 60 months.

(G) $R_0 = 167/801 = 0.02082$ (approx. 2.1%) R_1 23/456 = 0.05044 (approx. 5.0%) R_2 213/4231 = 0.05034 (approx. 5.0%), R_3 = 13/282 = 0.04610 (approx. 4.6%). The average risks of coronary events in low anger-temperament normotensive group (risk 2.1%) is less than half that of the other groups.

7.5 Depression and Parkinson's disease

(A) This is a prospective cohort study.

(B) Incidence proportion.

(C) The following measures of association are possible: incidence proportion ratio (the risk ratio), the incidence odds ratio (''odds ratio''), and the incidence proportion difference (risk difference).

(D) RR (215/1358)/(1845/57570) = 4.94. This indicates that the depressed subjects had about five times the risk of developing Parkinson's disease. It is also appropriate to say that depression is associated with approximately 400% greater risk of Parkinson's disease in relative terms.

Review questions

R.7.1 Rates from open population are not longitudinal because they do not normally permit the long-term follow-up of individual experience over time.

R.7.2 *Cohors* is Latin for ''an enclosure.'' It is also used to refer to a legion of men in the Roman army.

R.7.3 False. When an epidemiologist makes reference to a cohort study without specification they are probably referring to an observational cohort study.

R.7.4 Bernardino Ramazzini (1633–1714).

R.7.5 Percivall Pott (1713–1788) noted that chimney sweeps had enormously elevated rates of scrotal cancer and attributed this to chimney soot lodged in the rugae of their scrotum.

R.7.6 Goldberger observed that pellagra was common among the ''patient cohort'' at state hospitals but was absent among the ''staff cohort.'' He reasoned that if pellagra was communicable, it would not spare the staff cohort.

R.7.7 Wade Hampton Frost's generational cohort studies emulated the long-term follow-up of individuals needed to study the causes of chronic ''life style'' diseases.

R.7.8 Framingham is a moderately sized town in Massachusetts that became the seat of a cohort study that has since yielded many important findings about heart disease.

R.7.9 These are the two investigators who started The British Doctors study.

R.7.10 The British Doctors cohort has been followed since 1951 to present.

R.7.11 About 30%.

R.7.12 index

R.7.13 A cohort study is prospective if it accrues data in close proximity to the time of actual occurrence. Retrospective cohort studies use historical information.

R.7.14 Experimental cohort studies are always prospective because study subjects must be first be exposed to the study treatment before being followed for the outcome.

R.7.15 ''Historical cohort study'' is a synonym for ''retrospective cohort study.''

R.7.16 Confounding is the mixing of effects of the study exposure with extraneous factors, causing a false attribution (or lack of attribution) of effect.

R.7.17 Exposure groups are compared at the beginning of the follow-up period to assess group comparability and the potential for confounding.

R.7.18 Admissibility criteria can be used to restrict participants to individuals with or without certain characteristics that have the potential to confound the results.

R.7.19 Methods to mitigate confounding in observational cohort studies: (1) Statistical modeling or stratified analysis, (2) restriction through admissibility criteria, (3) matching.

Chapter 8: Case–control studies

Exercises

8.1 Influenza vaccination and primary cardiac arrest

(A) Cases were identified from the source population (King County, Washington). Noncases were selected from the community to serve as controls.

Exposure status (influenza vaccination) was ascertained in the case series and control series, making this a case–control study.

(B) Random digit dialing was used to derive a random sample of noncases from the source population to serve as controls.

(C) We speculate two reasons for using spouses of controls to provide exposure information. First, spouses are a good source of this type of information. Second, spouses of controls could be reasonably expected to derive information of comparable quality to that of cases. Thus, the information derived from controls would tend to be comparable accurate/inaccurate to that of cases, resulting in a balanced type of error.

(D) Two-by-two table:

	Cases	Controls
Vaccinated	79	176
Not vaccinated	236	373
	315	549

$$\widehat{OR} = \frac{79 \cdot 373}{236 \cdot 176} = 0.71$$

Interpretation: The vaccinated group had about 71% of the risk of the non-vaccinated group. That is, the vaccinated group had about 29% less risk.

Aside: The published article[a] reported an odds ratio of 0.51 adjusted for matching factors (age and gender), current smoking, former smoking, hypertension, diabetes, weight, height, habitual physical activity, saturated fat intake, family history of myocardial infarction or sudden death, educational attainment, employment, and general health status.

(E) The 95% confidence of the OR is (0.52 to 0.97) by the Mid-P exact method (calculated by *www.OpenEpi.com* version 2.3.1).

8.2 Wynder and Graham's pioneering study

(A) Lung cancer has a long induction. It is possible that many patients coming to the hospital with chronic lung disease would have stopped smoking months or years ago. It was therefore important to ask subjects about their prior smoking habits, at the time that lung cancer might have been induced.

(B) It is important to assess other factors that might be associated with the study outcome as possible determinants or potential risk factors for the current study.

(C) Confirmation of case status is important to avoid misclassification.

(D) $\widehat{OR}_0 = 1.0$ (referent), $\widehat{OR}_1 = 2.45$, $\widehat{OR}_2 = 5.97$, $\widehat{OR}_3 = 11.17$, $\widehat{OR}_4 = 27.28$, $\widehat{OR}_5 = 27.63$. Odds ratios show stronger and stronger associations with higher levels of smoking, indicating a dose–response relationship.

---

[a] Siscovick, D.S., Raghunathan, T.E., Lin, D., Weinmann, S., Arbogast, P., Lemaitre, R.N., *et al*. (2000). Influenza vaccination and the risk of primary cardiac arrest. *American Journal of Epidemiology*, **152** (7), 674–677.

8.3 Thrombotic stroke in young women.

(A) $\widehat{OR} = \frac{44}{5}$ = 8.8. There is an eightfold increase in risk among the oral contraceptive users.

(B) Exposed cases: 2 + 44 = 46.

(C) Nonexposed cases: 5 + 55 = 60.

(D) Exposed controls: 2 + 5 = 7.

(E) Nonexposed controls: 44 + 55 = 99.

(F) Match broken (tabulation below) OR = 10.8. This analysis inflated the odds ratio.

	Cases	Controls
OC +	46	7
OC −	60	99
	106	106

8.4 Esophageal cancer and alcohol consumption

(A) $\widehat{OR} = \frac{31 \cdot 447}{78 \cdot 51} = 3.48$.

(B) This level of tobacco consumption more than triples the risk of esophageal cancer.

(C) 95% confidence intervals = (2.08 to 5.77).

(D) This would bias the results away from the null, possibly producing a spurious association.

8.5 Pancreatic cancer and meat consumption

	Cases	Controls
E +	53	53
E −	46	85
	99	138

$\widehat{OR}$= (53 × 85) / (53 × 46) = 1.848 ≈ 1.8

Discuss the findings. Eating fried or grilled meat increased the risk of pancreatic cancer by about 80%.

8.6 IUDs and infertility (3 pts: 2-by-2 table, OR, interpretation)

	Cases	Controls
IUD +	89	640
IUD −	194	3193
	283	3833

$\widehat{OR} = (89 \times 3193)/(640 \times 194) = 2.29 \approx 2.3$

Interpretation: IUD users have 2.3 times the risk of infertility as nonusers. It is also appropriate to say use of these IUDs increased the risk of infertility by approximately 130%.

Review questions

R.8.1 The two most common types of analytic studies in observational epidemiology are cohort studies and case–control studies.

R.8.2 The main distinction between cohort studies and case–control studies is the way in which they select subjects for study. Cohort studies select healthy individuals and go on to determine the incidence of disease. Case–control studies select diseased individuals (cases) and non-diseased individuals (controls) and determine the occurrence of prior exposures.

R.8.3 Case–control studies cannot determine the incidence or prevalence of disease because they do not determine number of individuals or number of individuals at risk in the population or population subgroups.

R.8.4 Case–control studies can calculate odds ratio.

R.8.5 Relative risks (or rate ratios).

R.8.6 The subscript 1 denotes ''exposed'' while the subscript 0 denotes ''nonexposed.''

R.8.7 In this book's notation, A represents the number of cases and B represents the number of noncases in the study.

R.8.8 The odds ratio is occasionally referred to as the cross-product ratio.

R.8.9 An odds ratio of 1 represents no association between the exposure and disease.

R.8.10 An odds ratio of 2 represents a positive association in which exposed individuals have twice the risk of disease as the nonexposed individuals.

R.8.11 An odds ratio of 0.5 represents a negative association in which exposed individuals have half the risk as nonexposed individuals.

R.8.12 We can find cases for case–control studies in medical facilities, disease registries, surveillance systems, computerized medical record systems, and death certificates.

R.8.13 A case definition is the set of uniform criteria by which to decide whether an individual should be classified as a case in an epidemiologic investigation.

R.8.14 An incident case is one of onset during the period of observation. The onset of a prevalent case preceded the period of observation.

R.8.15 Controls may be derived from random samples from the population, for the capturement area from the clinics and hospitals that were the source of cases, from among hospitalized individuals with ailments that are unrelated to the disease being studied, and so on.

R.8.16 The primary function of the control series in a case–control study is to reflect the relative frequency of the exposure in the source population.

R.8.17 The best source of population-based controls for hospitalized cases is the capturement area for that hospital.

R.8.18 A nested case–control study identifies cases as part of a cohort study. It then samples non-cases from the study cohort to serve as controls.

R.8.19 Case–control studies are statistically efficient when studying rare diseases because they use a small fraction of the non-cases from the source population to estimate the frequency of exposure in the underlying population.

R.8.20 The correct answer is (c).

R.8.21 The correct answer is (c).

R.8.22 1.

R.8.23 4.

Chapter 9: Sources of error

Exercises

9.1 Biased surveys

(A) A certain percentage of persons who have had venereal disease may deny it or simply not know what the term ''venereal disease'' means. Thus, *information bias* is likely to occur; prevalence will likely be underestimated.

(B) Seniors with disabilities would be less likely to attend dance lessons. Therefore, a *selection* bias is likely; the true prevalence of disability will likely be underestimated.

(C) Some people may be unfamiliar with the term ''coronary artery disease.'' Therefore, an *information bias* may occur, possibly resulting in under-reporting.

(D) We can assume a certain percentage of teenagers who smoke will deny smoking, resulting in an *information bias* that will underestimate the true prevalence.

(E) We may hypothesize that workers without benefits would have a disincentive for missing work. Thus, the condition may well be under-reported as a form of *information bias*.

(F) People with dementia may no longer live at home. This survey would tend to underestimate the prevalence of dementia because of this undercoverage, which is a form of *selection bias*.

9.2 Confounding by race?

(A) Yes, the properties of confounding have been met. SES is associated with race, and race is a risk factor for hypertension. In addition, race is not intermediate in the causal pathway between SES and hypertension. Thus, race is likely to confound the association between socioeconomic status and hypertension.

(B) Race is a confounder. There is no association between SES and hypertension. However, race is associated with both hypertension and with socioeconomic status.

9.3 Meta-analysis of coronary heart disease treatment options

(A) The choice of procedure (CABG vs. PCI) has the potential to confound the results of observational studies because (a) treatment choice is associated with the severity of the underlying condition and (b) severity of the condition is an independent risk factor for death. (c) In addition, the severity of the condition is not intermediate in the causal pathway.

(B) With the conditions as stated, the bias would favor PCI relative to CABG because the more severe cases are being channeled to CABG.

(C) Randomization takes the treatment choice out of the hands of doctors and their patients and gives it to the flip of a coin. This tends to create treatment groups that are balanced with respect to measured and unmeasured confounders other than the treatment, thus breaking one of the key conditions for confounding.

(D) R_{CABG} = 574/3889 = 14.8%; R_{PCI}; 628/3923 = 16.0% The large p-value indicates that the slight advantage associated with PCI is consistent with a chance finding.

(E) There appears to be an advantage of CABG in older patients, possibly men, diabetics, and non-smokers. There is no clear advantage according to hypertension status, peripheral vascular disease status, or whether the patient experienced a previous heart attack.

9.4 Alternative explanations for high rates of disability in monks There are several alternative explanations for his finding. For example, Monks might be less hesitant to report disabilities than other comparably aged men. Alternatively, there could be a selection bias operative whereby men with physical disabilities may be more likely to choose the monastic life than men without physical disabilities.

Review questions

R.9.1 The two types of error in epidemiologic research are systematic error and random error.

R.9.2 Parameters can not be calculated and are assumed to be error free quantifications of the measurement in question. Estimates are calculated using empirical data and are prone to both random and systematic sources of error.

R.9.3 Bias is a synonym for systematic error.

R.9.4 Imprecision is a synonym for random error.

R.9.5 Valid is an antonym for biased.

R.9.6 Imprecise is an antonym for precise.

R.9.7 Random error is balanced, decreases with sample size, and can be dealt with via confidence intervals and p values. Systematic error is unbalanced, is not affected by sample size; routine statistical inferential methods such as confidence intervals and p values do not address systematic error.

R.9.8 False. Probability models address random error only.

R.9.9 Estimation and hypothesis (significance) testing.

R.9.10 Expanding the size of an observational study will decrease its random error and have no effect on its systematic error.

R.9.11 Selection bias, information bias, confounding bias.

R.9.12 Nondifferential misclassification biases measures of association either toward the null or not at all.

R.9.13 True. These are good definitions of bias toward the null.

R.9.14 Confounding is a distortion in a measure of association brought about by extraneous factors ''lurking'' in the background.

R.9.15 Properties of a confounder: (1) associated with the exposure, (2) independent risk factor for disease, (3) not intermediate in the causal pathway.

R.9.16 The Latin term *confundere* means ''to mix-up.''

R.9.17 Use of hospitalized cases and controls in a case–control study may result in hospital admission rate bias, also known as Berkson's bias.

R.9.18 This will cause the type of information bias referred to as recall bias. As described, this will bias the results of the study away from the null.

R.9.19 This will cause an information or misclassification bias.

R.9.20 A risk factor that is equally distributed in the groups being compared will not confound the results of the study. In addition, a risk factor that is intermediate in the causal pathway between the exposure and disease should not be considered to be a risk factor.

R.9.21 The results of a study of the effects of smoking on lung cancer done in men can be applied to women because the causal mechanisms of lung cancer are similar in men and women.

R.9.22 False. Confidence intervals address random error only and do not address systematic sources of error. Therefore, the confidence is limited to the consideration of the precision of the estimate and not to the overall validity.

Chapter 10: Screening for disease

Exercises

10.1 Sign, symptom, or test?

(A) Chills, as subjective sensations of the patient, are symptoms.
(B) Body temperature in degrees Fahrenheit is a test.
(C) Soreness, as reported by the patient, is a symptom.
(D) Swelling and redness as observed by a clinician is a sign.

10.2 Cross-tabulate first

(A) There are two false positive specimens: specimen 1 and specimen 9.
(B) There is one false negative specimen: specimen 10
(C) See table:

		Test A (Gold standard)		
		+	−	Total
Test B	+	3	2	5
Test B	−	1	4	5
		4	6	10

(D) SEN = TP/(TP + FN) = 3/4 = 0.750.
(E) SPEC = TN/(TN + FP) = 4/6 = 0.667.

10.3 Sensitivity, specificity and predictive value
(A) SEN = 40/50 = 0.800.
(B) SPEC = 125/150 = 0.833.
(C) PVPT = 40/65 = 0.615.
(D) PVNN = 125/135 = 0.926.

10.4 Use Bayesian formula

$$\begin{aligned}\text{PVPT} &= \frac{(P)(\text{SEN})}{(P)(\text{SEN}) + (1 - \text{SPEC})(1 - P)} \\ &= \frac{(0.001)(0.95)}{(0.001)(0.95) + (1 - 0.90)(1 - 0.001)} = 0.0094\end{aligned}$$

Thus, less than 1% of people testing positive on the screening test will actually have the cancer in question.

10.5 Drug test The problem provides a prevalence of 0.05, sensitivity of 0.95, and specificity of 0.95. The question requests the predictive value positive of a test. According to Formula (10.13),

$$\begin{aligned}\text{PVPT} &= \frac{(P)(\text{SEN})}{(P)(\text{SEN}) + (1 - \text{SPEC})(1 - P)} \\ &= \frac{(0.05)(0.95)}{(0.05)(0.95) + (1 - 0.95)(1 - 0.05)} = 0.50\end{aligned}$$

Therefore, only half of the people testing positive will actually be drug users.

10.6 Two raters

$$p_o = \frac{150 + 239}{448} = 0.8683$$

$$p_e = \frac{178 \cdot 181 + 270 \cdot 267}{448^2} = 0.5197$$

$\kappa = \dfrac{p_o - p_e}{1 - p_e} = \dfrac{0.8683 - 0.5197}{1 - 0.5197} = 0.7258$ This level of agreement is regarded as substantial.

10.7 Sensitivity and specificity

(A) $SEN = \dfrac{TP}{\text{those with disease}} = \dfrac{15}{18} = 0.8333$ or about 83%

$$SPEC = \frac{TN}{\text{those free of disease}} = \frac{145}{152} = 0.9539 \text{ or about } 95\%$$

(B) This exercise is a validity analysis because it gauges the results of a test against a definitive diagnosis (''gold standard''). The prior exercise was a reproducibility analysis because it considers agreement between two raters, neither of which is a gold standard.

10.8 Predicting tornados

(A) $\kappa = 0.56$. This kappa statistic demonstrates that Finley is doing better than chance. Supplemental analysis: The p-value for testing H_0: $\kappa = 0$ is less than 0.0005 (calculated with WinPEPI → PAIRSetc → A. 'Yes-no' variable. This confirms that Finley is not merely guessing.

(B) SEN = 11/14 = 78.6%. Note that Finley missed 3 of the 14 (21.4%) tornados.

(C) PVPT = 11/25 = 0.4400, indicating that Finley was correct only 44% of time when predicting a tornado. The PVNT = 906/909 or 99.7%, which seems pretty good, but he did miss 3 occurrences.

10.9 Screening for bladder cancer

(A) This is a comparison of diagnoses from two raters, neither of which is a gold standard.

(B) $\kappa = 0.898$, indicating a high level of agreement.

(C) See table:

	Disease		
Test	+	−	Total
+	90	1998	2088
−	10	97 902	97 912
Total	100	99 900	100 000

(D) PVPT = 90/2088 = 0.0431.

(E) PVNT = 97 902/97 912 = 0.9999.

(F) Table:

	Disease		
Test	+	−	Total
+	90	18	108
−	10	882	892
Total	100	900	1000

(G) PVPT = 90/108 = 0.8333

(H) Because the prior probability (prevalence) of disease was much higher in this second group.

10.10 Leg length inequality

p_{obs} = (12 + 28)/46 = 0.87; κ = 0.70 (indicating substantial reliability); p_{pos} = 0.800 and p_{neg} = 0.903 (indicating that reliability was comparable with identifying either short leg).

Review questions

R.10.1 Reliability refers to repeatability or consistency of a result from one use to the next. Validity is the ability to discriminate between those with and with a condition.

R.10.2 Symptom = experienced by a patient; Sign = observation made by an examiner.

R.10.3 Reliability statistics: κ, proportion of specific positive agreement (p_{pos}), and proportion of specific negative agreement (p_{neg}). Validity: SEN, SPEC, PVPT, and PVNT.

R.10.4 κ takes chance agreement into account.

R.10.5 False. A κ of 0 indicates no further agreement above that expected from chance.

R.10.6 (a) A TP is someone with disease that tests positive. (b) A TN is someone without disease who tests negatively. (c) An FP is someone without disease who has a positive test. (d) An FN is someone with disease who tests negatively.

R.10.7 Matching: (a) Specificity = Pr(T−|D−), (b) Sensitivity = Pr(T+|D+), (c) PVPT = Pr(D+|T+), and (d) PVNT = Pr(D−|T−).

R.10.8 ''Specificity'' is the expected proportion of disease-free individuals that will have a negative test.

R.10.9 The ''predictive value of a negative test'' is the proportion of negative tests that are disease-free.

R.10.10 PVPT is determined by the prevalence of disease, sensitivity of the test, and specificity of the test.

R.10.11 The PVPT will tend to be low when the prevalence of disease is low because even a specific test will identify many false positives under such circumstance. For example, a test that is 99.9% specific when used in a million disease-free people will derive 0.001 × 1 000 000 = 1000 false positives.

R.10.12 It will decrease sensitivity, but will increase specificity.

R.10.13 It should be of high sensitivity, since you don't want to miss cases and the consequences of false positives are minimal.

Chapter 11: The infectious disease process

Exercises

11.1 (A) A portal is an entry or exit site for the agent. Agents cannot enter directly through the cardiovascular system.

11.2 (C) Antibodies are not innate. They are acquired as a result of exposure to a specific antigen.

11.3 (C) Toxoids invoke active responses to specific toxins.

11.4 (D) Maternal transfers are a form of passive immunization.

11.5 (H) All of the items listed may serve as reservoirs.

11.6 (C) Incomplete treatment with antibiotics causes the increased likelihood of the convalescent carrier state.

11.7 (B) Inapparent carriers remain disease-free throughout infection. Incubatory carriers are yet to demonstrate signs of disease. Chronic and convalescent carriers have recovered from the disease in question.

11.8 (D) Animate transmitters of infectious agents are called vectors.

11.9 (A) By definition, direct zoonoses have a single animal reservoir. Cyclozoonoses require two vertebrate species to complete their life cycle; metazoonoses require an invertebrate host or vector in addition to the vertebrate reservoir; and saprozoonoses require an inanimate reservoir in addition to their animal reservoirs.

Review questions

R.11.1 Contamination is the presence of the agent on a surface. Infection is presence of the agen within the body of the host where it is capable of multiplying.

R.11.2 Infections may or may not have signs and symptoms. Infectious diseases are associated with signs, symptoms, and physiologic dysfunction.

R.11.3 Helminth, fungi and yeast, protozoans, bacteria, rickettsia, viruses, prions.

R.11.4 Types of reservoirs: Cases, carriers, animals, and inanimate objects

R.11.5 Cases demonstrate signs or symptoms; carriers do not.

R.11.6 Zoonotic disease.

R.11.7 Water, soil, food, air.

R.11.8 Cyclozoonoses require two vertebrate hosts to complete their cycle. Direct zoonoses travel directly from a non-human animal to a human.

R.11.9 Metazoonoses require invertebrate intermediaries (e.g., insects) between vertebrate species.

R.11.10 ''Sapro'' means ''dead.''

R.11.11 Portals: upper respiratory, conjunctival, urogenital, gastrointestinal, skin, placenta.

R.11.12 Direct contact transmission requires physical contact between a contagious and susceptible host. Indirect contact requires contact between infectious material and the susceptible host.

R.11.13 Droplets → large infectious particles transmitted via spray. Droplet nuclei → small aerosolized particles suspended in air.

R.11.14 A vector is a living animal or insect. A vehicle is inanimate.

R.11.15 Mechanical transmission → no multiplication of agent in vector or vehicle. Developmental transmission → agent undergoes biological transformation or maturation in vector or vehicle.

R.11.16 With developmental transmission, the agent undergoes a biological transformation in the vehicle or vector. With propagative transmission, the agent multiplies in the vehicle or vector without going through a biological transformation.

R.11.17 The Broad Street pump outbreak was a common-source outbreak.

R.11.18 It is believed that the common cold spreads serially, from person-to-person.

R.11.19 Understanding the biological cycle of an agent permits multiple opportunities to disrupt the ecology of the disease.

R.11.20 Understanding the natural history of a disease in an individual may shed light on periods of communicability and susceptibility. It may also help with understanding routes of transmission and portals for infection.

R.11.21 True.

R.11.22 Active immunization requires an immune response on the part of the host. Passive immunity does not.

R.11.23 Two forms of passive immunization: therapeutic (e.g., immunoglobin injection) and maternal transfer (transplacentally or through colostrum).

R.11.24 vaccination.

R.11.25 Modified live vaccination generally elicit the more sustained immune responses since the host is exposed to the antigen challenge for a longer period of time.

R.11.26 Cytokines and chemokines.

R.11.27 B lymphocytes.

R.11.28 T4 helper cells.

R.11.29 lymphocytes; a type of white blood cell.

R.11.30 (1) High incidence of infectious diseases internationally, and emerging infectious diseases. (2) May provide insights into epidemiology in general (concept of ''one epidemiology'').

R.11.31 Herd immunity is an effect beyond individual immunity when a certain proportion of the population is immune. Innate herd immunity encompasses inborn (innate) forms of immunity.

R.11.32 Acquired herd immunity requires exposure to the agent and an active (physiologic) response on part of the herd (population) members.

R.11.33 When a high percentage of individuals in the ''herd'' are immune, transmission finds dead ends, preventing further spread.

Chapter 12: Outbreak investigation

Review questions

R.12.1 The Centers for Disease Control and Prevention (CDC).

R.12.2 (b)

R.12.3 (d)

R.12.4 (a)

R.12.5 Outbreaks are detected by surveillance systems or concerned citizens and health professionals.

R.12.6 Surveillance systems are organizations and structures set up to routinely collect outcome-specific health data.

R.12.7 A change in the reporting procedure or case definition; an increase in population size; diagnostic suspicion bias; publicity bias.

R.12.8 (a) The ability to confirm a greater than expected number of cases truly occurred, (b) the scale and severity of the outbreak, (c) whether a identifiable subgroup is disproportionally affected, (d) the potential for spread, (e) political and public relations considerations, (f) availability of resources.

R.12.9 Person, place, and time factors.

R.12.10 Diagnoses.

R.12.11 Expected.

R.12.12 A case definition is the standard criteria used to decide whether someone has or does not have the disease in question.

R.12.13 True.

R.12.14 The incubation period is the time interval between exposure to the agent and the appearance of first signs or symptoms.

R.12.15 A hypothesis is a tentative explanation that accounts for a set of facts that can be tested by further investigation.

R.12.16 True.

R.12.17 how and why.

Chapter 14: Mantel–Haenszel methods

Exercise

14.1 (A) The proportion of children with health-care coverage in the traced group = 46/370 = 11%. The proportion with health-care coverage in the lost-to-follow-up group = 195/1174 = 17%. Therefore, there is a negative association between being traced and health-care coverage.

(B) Among whites, 83% of those not lost-to-follow-up had health-care coverage (10/12). The same percentage holds for whites that were not lost-to-follow-up (104/126). Among blacks, 36/404 (9%) of those not lost-to-follow-up had health-care coverage. The same percentage holds for blacks that were lost-to-follow-up (91/1048). Therefore, there is no association between being traced and health-care coverage within race groups.

(C) This is an example of Simpson's paradox. The negative association between those remaining in the cohort and health-care coverage was eliminated upon stratification by race. An explanation for this is that whites tended to

have more health-care coverage and were also more likely to be lost-to-follow-up. Therefore, the comparison between being traced and health-care coverage was to some extent a comparison of health-care coverage by race.

(D) The unconfounded incidence proportion ratio associated with being traced/lost-to-follow-up after controlling for race should be about 1.

Chapter 15: Statistical interaction

Exercises

15.1 Nischan *et al.* (1988)

(A) $\hat{\psi}_1 = 2.7$

(B) $\hat{\psi}_2 = 1.1$

(C) H_0: $\psi_1 = \psi_2$;
$\chi_{INT}^2 = (0.9922 - 0.2311)^2/(0.4236)^2 + (0.0974 - 0.2311)^2/(0.1871)^2 = 3.80$
$df = 2 - 1 = 1$, $p = 0.051$

(D) Based on the discrepant point estimates and test of statistical interaction, I would conclude that statistical interaction is present. Therefore, I would *not* report a Mantel–Haenszel summary odds ratio in this situation. Instead, I would report separate strata-specific odds ratios.

15.2 Colditz *et al.* (1990)

(A) $\hat{\omega}_1 = 1.05$

(B) $\hat{\omega}_2 = 1.65$

(C) H_0: $\omega_1 = \omega_2$, $\chi_{INT}^2 = (0.0497 - 0.2791)^2/(0.1342)^2 = (0.5014 - 0.2791)^2/(0.1252)^2 = 6.07$; $df = 2 - 1 = 1$, $p = 0$ .014. This, along with the different strata-specific estimates, suggests that statistical interaction (effect measure heterogeneity) is present.

Chapter 17: Survival analysis

Exercises

17.1 Stanford Heart Transplantation Program data

(A) See Table 17.7 on page 371.

(B) See Figure 17.7 on page 432.

(C) Based on the Figure, the median survival is approx. 250 days.

17.2 The cumulative survival in the treatment group is 0.5714, and the cumulative survival in the control group is 0.2400. Therefore, $\hat{P}_{1,96\ \text{mo.}} = 1 - 0.5714 = 0.4286$, $\hat{P}_{0,96\ \text{mo.}} = 1 - 0.2400 = 0.7600$, and $\widehat{\phi}_{96\ \text{mo.}} = 0.4286/0.7600 = 0.56$. Eight-year survival in the control group is about half that of the treatment group.

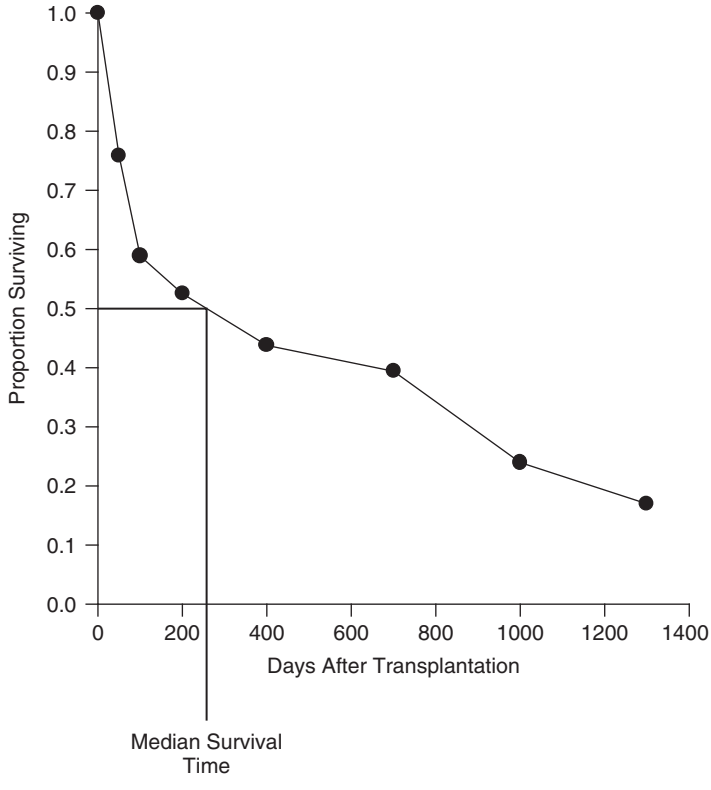


Figure 17.7 Survival following heart transplantation; answer to Example 17.1B.

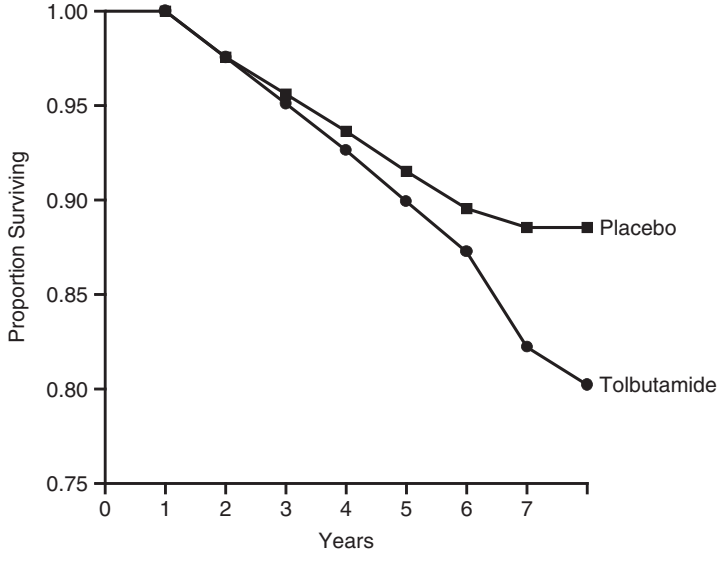


Figure 17.8 Survival, tolbutamide vs. placebo; answer to exercise 17.3(A). (Source: UGDP, 1970).

17.3 UGDP: tolbutamide vs. placebo

(A) See Figure 17.8 on the prior page. The placebo group demonstrated superior survival.

(B) The Cochran–Mantel–Haenszel risk difference is .0077 ($\chi^2_{CMH} = 1.82$, df = 1, $p = 0.18$). Thus, there was an excess mortality of 0.77% per year with tolbutamide. However, according to this analysis, this excess could possibly be explained by chance ($p = 0.18$).

Chapter 18: Current life tables

Exercises

18.1 No answer provided.

18.2 See Table 18.5.

Table 18.5 Completed Table for Exercise 18.2

Age range x to $x + n$	Proportion dying ${}_nq_x$	No. living at beginning of age interval l_x	No. dying during the age interval ${}_nd_x$	Expected person-years in age interval ${}_nL_x$	Expected person-years in this and all subsequent age intervals T_x	Expected no. of years of life remaining at beginning interval of age e_x
0–1	0.06392	100 000	6392	94 247	5 851 258	58.5
1–5	0.03084	93 608	2887	365 194	5 757 010	61.5
5–10	0.01275	90 721	1157	450 713	5 391 817	59.4
10–15	0.00951	89 564	852	445 690	4 941 104	55.2
15–20	0.01613	88 712	1431	439 983	4 495 414	50.7
20–25	0.02427	87 281	2118	431 110	4 055 432	46.5
25–30	0.02845	85 163	2423	419 758	3 624 322	42.6
30–35	0.03063	82 740	2534	407 365	3 204 564	38.7
35–40	0.03219	80 206	2582	394 575	2 797 199	34.9
40–45	0.03547	77 624	2753	381 238	2 402 624	31.0
45–50	0.04440	74 871	3324	366 045	2 021 387	27.0
50–55	0.05904	71 547	4224	347 175	1 655 342	23.1
55–60	0.08346	67 323	5619	322 568	1 308 167	19.4
60–65	0.12001	61 704	7405	290 008	985 599	16.0
65–70	0.17792	54 299	9661	247 343	695 592	12.8
70–75	0.26572	44 638	11 861	193 538	448 249	10.0
75–80	0.37481	32 777	12 285	133 173	254 712	7.8
80–85	0.51645	20 492	10 583	76 003	121 539	5.9
85–90	0.65970	9909	6537	33 203	45 537	4.6
90–95	0.78618	3372	2651	10 233	12 334	3.7
95–100	0.91262	721	658	1960	2102	2.9
100+	1.00000	63	63	142	142	2.25

Chapter 19: Random distribution of cases in time and space

Exercises

19.1 **(A)** $\Pr(A = 0) = (e^{-1.2}\ 1^0)/(0!) = 0.301\,194$ (by Formula 19.1).
(B) $\Pr(A = 1) = 0.361\,433$
(C) $\Pr(A = 2) = 0.216\,860$
(D) $\Pr(A \geq 3) = 1 - \Pr(A \leq 2) = 1 - (3012 + 0.3614 + 0.2169) = 0.1205$
(E) Not very. This would occur about 12% of the time.

19.2 **(A)** 0.1353
(B) 0.2706
(C) 0.2706
(D) 0.1804
(E) 0.1431

19.3 **(A)** $\mu = (n)(\lambda_0) = (7076)(286/1\,152\,695) = 1.76$.
(B) $\Pr(A \geq 8)$ under these assumptions = 0.00049. Therefore, this does represent an unusually high number of cases. However, there is still a small probability (1 in 2000) that the observation represents a chance occurrence. In addition, the geographic boundaries for the cluster were drawn after the fact, making this akin to shooting an arrow and drawing the bull's-eye after that arrow has landed. This demonstrates one of the difficulties in interpreting the post hoc identification of clusters.

19.4 $\Pr(A \geq 4) = 0.1087$.

19.5 **(A)** 0.4724
(B) 0.3543
(C) 0.1329
(D) 0.0332
(E) $\Pr(A \geq 4) = 1 - (0.4724 + 0.3543 + 0.1329 + 0.0332) = 0.0072$.

APPENDIX 1

95% Confidence Limits for Poisson Counts

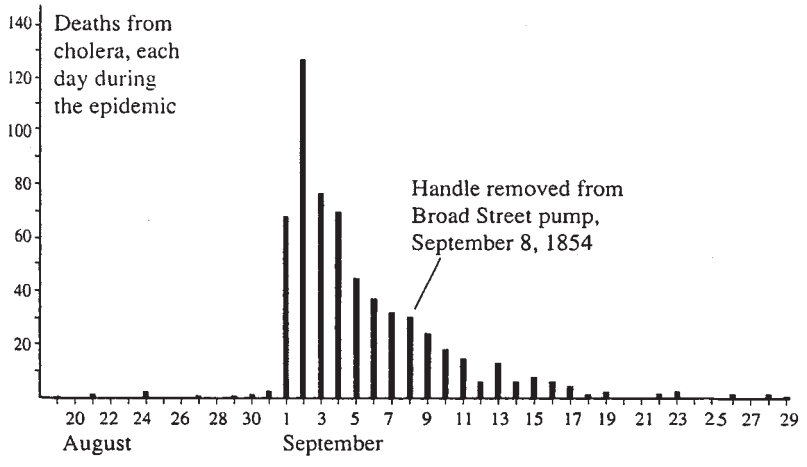
Observed number of cases	A_{LCL}	A_{UCL}
0	0.0000	3.69
1	0.0253	5.57
2	0.242	7.22
3	0.619	8.77
4	1.09	10.24
5	1.62	11.67
6	2.20	13.06
7	2.81	14.42
8	3.45	15.76
9	4.12	17.08
10	4.80	18.39
11	5.49	19.68
12	6.20	20.96
13	6.92	22.23
14	7.65	23.49
15	8.40	24.74
16	9.15	25.98
17	9.90	27.22
18	10.67	28.45
19	11.44	29.67
20	12.22	30.89
21	13.00	32.10

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

Observed number of cases	A_{LCL}	A_{UCL}
22	13.79	33.31
23	14.58	34.51
24	15.38	35.71
25	16.18	36.90
26	16.98	38.10
27	17.79	39.28
28	18.61	40.47
29	19.42	41.65
30	20.24	42.83
31	21.06	44.00
32	21.89	45.17
33	22.72	46.34
34	23.55	47.51
35	24.38	48.68
36	25.21	49.84
37	26.06	51.00
38	28.46	52.15
39	27.73	53.31
40	28.58	54.47

APPENDIX 2

Tail Areas in the Standard Normal (Z) Distribution: Double These Areas for Two-Sided p-Values



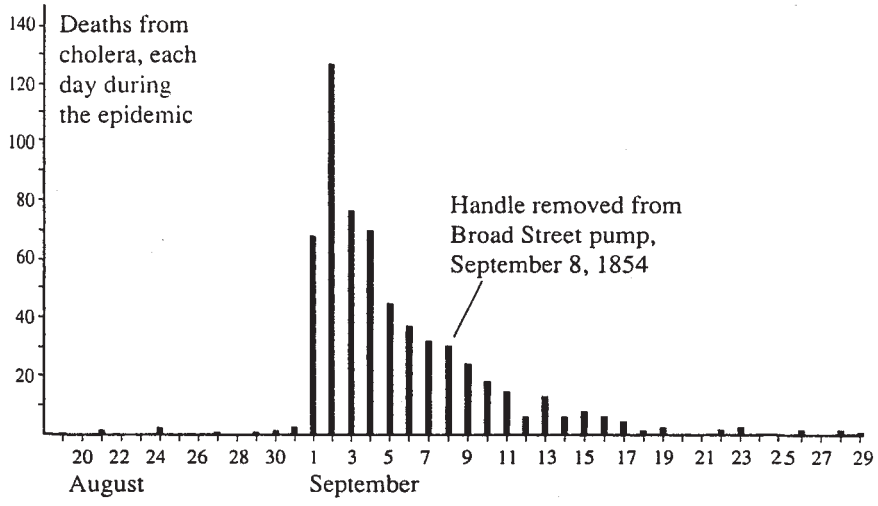
Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

Z Values in Tenths	*Z* Value in Hundredths									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.10	0.46017	0.45620	0.45524	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.20	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
0.30	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
0.40	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
0.50	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
0.60	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
0.70	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
0.80	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
0.90	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
1.00	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
1.10	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
1.20	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
1.30	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
1.40	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
1.50	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
1.60	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
1.70	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
1.80	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
1.90	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	*0.02500*	0.02442	0.02385	0.02330
2.00	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.10	0.01786	0.01743	0.01700	0.01659	0.01618	0.0578	0.01539	0.01500	0.01463	0.01426
2.20	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101

	Z Value in Hundredths									
Z Values in Tenths	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.30	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
2.40	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
2.50	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
2.60	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
2.70	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
2.80	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
2.90	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
3.00	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
3.10	0.00097	0.00094	0.00090	0.00087	0.00084	0.0082	0.00079	0.00076	0.00074	0.00071
3.20	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
3.30	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
3.40	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024

APPENDIX 3

Right-Tail Areas in Chi-Square Distributions



	Right-tail region									
df	0.995	0.975	0.95	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.000	0.001	0.004	1.32	2.71	*3.84*	5.02	6.63	7.88	10.83
2	0.010	0.051	0.103	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	0.072	0.216	0.352	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	0.207	0.484	0.711	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	0.41	0.83	1.15	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	0.68	1.24	1.64	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	0.99	1.69	2.17	9.04	12.02	14.07	16.01	18.48	20.28	24.32

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

	Right-Tail Region									
df	0.995	0.975	0.95	0.25	0.1	0.05	0.025	0.01	0.005	0.001
8	1.34	2.18	2.73	10.22	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.70	3.33	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	3.25	3.94	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.82	4.57	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	3.07	4.40	5.23	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	5.01	5.89	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	5.63	6.57	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	6.26	7.26	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	6.91	7.96	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	7.56	8.67	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	8.23	9.39	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	8.91	10.12	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	9.59	10.85	23.83	28.41	31.41	34.17	37.57	40.00	45.31
21	8.03	10.28	11.59	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	10.98	12.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	11.69	13.09	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	12.40	13.85	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	10.52	13.12	14.61	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	11.16	13.84	15.38	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	11.81	14.57	16.15	31.53	36.74	40.11	43.19	46.96	49.64	55.48

APPENDIX 4

Case Study—Cigarette Smoking and Lung Cancer

This case study is based on Centers for Disease Control and Prevention Epidemic Intelligence Service Summer Training Course, 1992. Concepts, terminology, and notation have been modified to be consistent with this text and materials have been supplemented.

Objectives

After completing this case study, the student should be able to:

1. Discuss the elements of case–control and cohort study designs and identify advantages and disadvantages to both.
2. Discuss some of the biases that affect epidemiologic studies.
3. Calculate and interpret odds ratios, rate ratios, rate differences, and attributable fractions.
4. Appreciate how the above measures of association do or do not reflect strength of association and public health importance.
5. Discuss causal criteria presented by Hill (1965).

Part I

A causal relation between cigarette smoking and lung cancer was first suspected in the 1920s on the basis of clinical observations. For many years it was debated whether the increase was real or an artifactual increase due to improved diagnostics. (The lungs were relatively inaccessible for diagnosis back then.) By the 1940s, there was some agreement on the veracity of the increase in occurrence, and the focus shifted to cause. Many theories were entertained, one of which was the smoking theory. (Other theories that were being entertained were the emissions from gas works and the fumes from the tarring of roads.) Two early studies on smoking and lung cancer were Lombard and Doering (1928) and Müller (1939). Some two decades later, two

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

studies appeared in *JAMA* supporting the association between tobacco smoking and lung cancer (Wynder and Graham, 1950; Levin *et al*., 1950; see Section 11.5). This case study focuses on several British studies completed between 1950 and 1964 by the team of Doll and Hill.

The first of these studies is a case–control study begun in 1947 in which smoking habits of lung cancer patients were compared with the smoking habits of other patients (Doll and Hill, 1950, 1952). The second study is a cohort study begun in 1951 in which the recorded causes of deaths among British physicians were studied in relation to smoking habits.

Data for the case—control study were obtained from hospitalized patients in London and vicinity over a 4-year period (April 1948 to February 1952). Initially, 20 hospitals, and later more, were asked to notify the investigators of all patients admitted with a new diagnosis of lung cancer. These patients were then interviewed concerning smoking habits, as were controls selected from patients with other disorders (primarily nonmalignant) hospitalized in the same hospitals at the same time.

Data for the cohort study were obtained from physicians listed in the *British Medical Register* who resided in England and Wales as of October 1951. Information about present and past smoking habits was obtained by questionnaire. Information about lung cancer came from death certificates and other mortality data recorded during the ensuing years (Doll and Hill, 1954, 1964).

Question 1a What makes the first study a case—control study?
Question 1b What makes the second study a cohort study?

The remainder of Part I deals with the case—control study.

Question 2 Why were hospitals chosen as the setting for this study? What other sources of cases and controls might have been used?
Question 3 What are the advantages of selecting controls from the same hospital as cases?
Question 4a The case series are all patients admitted to some 20-odd hospitals with a new diagnosis of lung cancer. How would you define the study base (''source population'') for these cases?
Question 4b The controls were patients with other disorders treated at these same hospitals. Do you think the controls would fairly represent the study base?
Question 4c How may these issues of representativeness affect the study's results?

More than 1700 cases of lung cancer all under age 75 were identified as potential cases. About 15% were not interviewed because of death, discharge, severity of illness, or inability to speak English. An additional group of patients were interviewed but later excluded when it was determined that the initial lung cancer diagnosis proved mistaken. The final group of cases consisted of 1357 males and 108 females (total, 1465). We will restrict future analyses to the male cases.

Table A4.1 is a 2-by-2 cross-tabulation of cigarette smoking in cases and controls.

Question 5 Calculate the odds ratio of lung cancer associated with smoking. Include a 95% confidence interval for the odds ratio. You may use *EpiCalc2000* or any

Table A4.1 Cigarette smoking and lung cancer, case—control study.

	Cases	Controls
Smoke +	1350	1296
Smoke –	17	61
	1357	1357

Table A4.2 Daily consumption of cigarettes in cases and controls.

Cigarettes per day	Cases	Controls
25+	340	182
15–24	445	408
1–14	565	706
0	7	61
	1357	1357

other epidemiologic calculator to compute the confidence interval. Interpret your results.

Table A4.2 shows the frequency of amount smoked in cases and controls.

Question 6 Calculate the odds ratio associated with each level of smoking compared to the baseline provided by nonsmokers. Interpret your results.

Question 7 While this study demonstrates a clear association between smoking and lung cancer, cause and effect are not the only possible explanation. What are other possible explanations for the association?

Part II

Part II of this exercise will address the cohort study. Data were obtained from the cohort of physicians listed in the *British Medical Register* who resided in England and Wales. Questionnaires were mailed in October 1951 to 59 600 physicians. The questionnaire asked the physicians to classify themselves into one of three smoking categories: (1) current smoker, (2) ex-smoker, or (3) nonsmoker. Smokers and ex-smokers were asked the amount they had smoked, their method of smoking, the age they started to smoke, and, if they had stopped smoking, how long it had been since they last smoked. Nonsmokers were defined as persons who had never consistently smoked as much as one cigarette a day for as long as one year.

Usable responses were received from 40 687 (68%) of the physicians, of which 34 445 were male and 6192 were female.

Question 8 How might the response rate of 68% affect the study's results?

Table A4.3 Lung cancer mortality according to smoking status.

	Lung cancer cases	Person-years
Cigarette smokers	133	102 000
Nonsmokers	3	42 800
All	136	144 800

The remainder of this case study addresses results in male physicians 35 years of age and older.

The lung cancer mortality in physicians responding to the questionnaire was studied over the 10-year period from November 1951 to October 1961 using information from the Registrar General Office and lists of death provided by the British Medical Association. Medical records were used to confirm the diagnosis. Seventy percent of cases were confirmed by biopsy, autopsy, and sputum cytology combined with bronchoscopy and X-ray evidence. Twenty-nine percent of cases were confirmed by cytology, bronchoscopy, or X-ray evidence alone. Approximately 1% of the diagnoses were based on case history, physical examination, or death certificate data only. Of the 4597 deaths in the cohort over the 10-year period, 157 were reported to be due to lung cancer. The diagnosis in four of these cases could not be confirmed, leaving 153 cases for study.

Question 9a Table A4.3 shows the number of lung cancer deaths and person-years in smokers and nonsmokers. Calculate the lung cancer rate in each group. Then calculate the rate ratio and rate difference associated with smoking. Interpret your results.

Question 9b Calculate the lung cancer rate in the entire cohort (smokers and nonsmokers combined). If no one had smoked in the cohort, we may assume that the cohort would have had the lung cancer rate of nonsmokers. What proportion of cases in the cohort would have been averted if no one had smoked? What is this fraction called?

Question 10 Table A4.4 lists the number of lung cancer deaths by amount smoked. Compute the rate ratio for each smoking category using the nonsmokers as the baseline rate in each instance. Interpret your results.

Table A4.4 Lung cancer mortality according to amount smoked.

Cigarettes per day	Lung cancer cases	Person-years
25+	57	25 100
15–24	54	38 900
1–14	22	38 600
0	3	42 800

Table A4.5 Lung cancer mortality in current smokers, former smokers, and nonsmokers.

Smoking status	Rate (per 1000 person-years)	Rate ratio
Current smokers	1.30	18.5
Former smokers, years since quitting		
<5	0.67	9.6
5 – 9	0.49	7.0
10 – 19	0.18	2.6
20+	0.19	2.7
Nonsmokers	0.07	1.0 (referent)

Question 11 Table A4.5 lists lung cancer mortality rates according to the duration of smoking cessation in people who had quit smoking. What do these data say about smoking cessation?

Question 12a The cohort study also provided information about cardiovascular mortality rates. Table A4.6 presents some of this data alongside lung cancer mortality data. Which disease, cardiovascular disease or lung cancer, has a stronger association with smoking?

Question 12b If the rate difference is used as an index of the effect of smoking per 1000 person-years of exposure, on which disease, cardiovascular disease or lung cancer, does smoking have a greater effect?

Question 13 Odds ratios from the case–control study and rate ratios from the cohort study are listed side by side in Table A4.7. How do these results compare? Can you suggest a plausible explanation for the higher relative risks in the cohort study?

Question 14a Using Table A4.8, check the column corresponding to the advantage held by the case–control or cohort study. For example, since case–control studies require smaller sample sizes than cohort studies, a check appears under the case–control column.

Question 14b Why was the case–control study done before the cohort study?

Table A4.6 Lung cancer and cardiovascular disease mortality in smokers and nonsmokers.

	Rate per 1000 person-years (all)	Rate per 1000 person-years (nonsmokers)	Rate per 1000 person-years (smokers)	Rate ratio	Rate difference per 1000 person-years	Population attributable fraction
Lung cancer	0.94	0.07	1.30	18.5	1.23	93%
Cardiovascular disease	8.87	7.32	9.51	1.3	2.19	17%

Table A4.7 Rate ratios and odds ratios according to amount smoked.

Daily number of cigarettes smoked	Rate ratio (cohort study)	Odds ratio (case—control study)
0	1.0 (reference)	1.0 (reference)
1–14	8.1	7.0
15–24	19.8	9.5
25+	32.4	16.3
All smokers combined	18.5	9.1

Table A4.8 Advantages according to study design (check advantage).

	Case–control	Cohort
Allows for smaller sample size	✓	
Costs less		
Shorter study time		
Better suited to study rare disease		
Better suited to study rare exposure		
Convenient when studying multiple exposures		
Convenient when studying multiple diseases		
Able to study natural history of disease		
Can estimate disease rates		
Less prone to selection biases		
Less prone to recall biases		
Less prone to loss to follow-up biases		

Question 15 In your opinion, which of the criteria for causality have been met by evidence presented in *this* case study?

Strength of association: (Y/N)
Consistency between studies: (Y/N)
Temporal sequence: (Y/N)
Biological gradient: (Y/N)
Specificity of effect: (Y/N)
Biological plausibility: (Y/N)

References

Doll, R., and Hill, A. B. (1950) Smoking and carcinoma of the lung. *British Medical Journal*, **2**, 739–748.

Doll, R., and Hill, A. B. (1952). A study of the aetiology of carcinoma of the lung. *British Medical Journal*, **ii**, 1271–1286.

Doll, R., and Hill, A. B. (1954). The mortality of doctors in relation to their smoking habits. *British Medical Journal*, **4877**, 1451–1455.

Doll, R., and Hill, A. B. (1964). Mortality in relation to smoking: 10 years' observation of British doctors. *British Medical Journal, I*, **1399–1410**, 1460–1467.

Hill, A.B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, **58**, 295–300.

Levin, M.L., Goldstein, H., and Gerhardt, P.R. (1950) Cancer and tobacco smoking. *JAMA*, **143**(*I*), 336–338.

Lombard, H.L., and Doering, C.R. (1928). Cancer studies in Massachusetts: Habits, characteristics and environment of individuals with and without lung cancer. *New England Journal of Medicine*, **198**, 481–487.

Müller, F.H. (1939). Tabakmissbrauch and Lungencazinom. *Zeitschrift fur Krebsforsuch*, **49**, 57–85.

Wynder, E.L., and Graham, E.A. (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *JAMA*, **143***(I)*, 329–336.

APPENDIX 5

Case Study—Tampons and Toxic Shock Syndrome

This case study is based on a Centers for Disease Control (CDC), Epidemic Intelligence Service 1992 Summer Training Course module. I have modified it to conform to the notation and terminology used in this book.

Objectives

After completing this case study, the student should be able to:

1. Describe the concepts, applications, and limitations of case–control studies
2. Analyze unmatched and matched case–control data
3. Discuss the valid selection of controls in case–control studies
4. Review sources of bias in case–control studies
5. Consider whether an association is causal

In 1979, three cases of an unusual illness were reported to the Wisconsin State Health Department. The three cases, all of which occurred in women, were characterized by fever, hypotension, diffuse rash, desquamation, and impairment of multiple organ systems. This clinical presentation was reminiscent of a recently described illness by Todd *et al*. (1978) given the name toxic shock syndrome (TSS). Todd's case series consisted of four girls and three boys 8–17 years old, five of whom had focal *Staphylococcus aureus* infections.

As a result of these case reports, Wisconsin and Minnesota established TSS surveillance systems within their states. By January of 1980, the two states had identified 12 cases, all in women. Eleven of the 12 women had been menstruating at the onset of the illness, and, anecdotally, ''most'' had been using tampons during the corresponding menstrual period. Soon thereafter, CDC was notified. In February, Utah established a TSS surveillance system.

During the spring of 1980, reports of TSS continued to trickle into CDC, mostly from Wisconsin, Minnesota, and Utah. The lead article of the May 23, 1980 issue

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.
© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

of *Morbidity and Mortality Weekly Report (MMWR)* described the first 55 cases of TSS reported to CDC. Of 40 patients in whom a menstrual history was obtained, 38 (95%) had onset of illness within 5 days following onset of menses. The case-fatality rate was 7/55, or 13%. In contrast, the case-fatality rate was 3.2% in Wisconsin, where surveillance had been proactive.

Extensive publicity followed the *MMWR* article, and CDC began to receive reports of TSS from other areas of the country.

Question 1 What biases would you be concerned about with this method of surveillance? In mid-June 1980, CDC conducted its first TSS case–control study (CDC-1). Published in the *MMWR* of June 27, 1980, this study of 52 female cases and 52 age- and sex-matched friend controls found a statistically significant association between tampon use and TSS. The report cited two separate studies from Wisconsin (31 cases) and Utah (12 cases). Data from these studies are shown in Tables A5.1 to A5.3.

Question 2 Using the small sample size odds ratio Formula (formula 11.3, p. 215), calculate the odds ratios for each of the case–control studies in Tables A5.1 to A5.3. From these data, would you conclude that TSS is associated with tampon use? Do you consider the Utah study to be consistent or inconsistent with the other two studies?

Table A5.1 Data from CDC-1 study[a].

	Cases	Controls
Tampon use +	50	43
Tampon use –	0	7

[a] Analysis of the 50 cases with onset during menstruation (MMWR, 27 June 1980).

Table A5.2 Data from the Wisconsin study.

	Cases	Controls
Tampon use +	30	71
Tampon use –	1	22

(*Source*: Davis *et al*. 1980)

Table A5.3 Data from Utah study.

	Cases	Controls
Tampon use +	12	32
Tampon use –	0	8

Table A5.4 Continual Tampon use during menstrual period among tampon users, case–control study of toxic shock syndrome, matched-pair analysis.

	Control exposed	Control nonexposed	
Case exposed	33	16	49
Case nonexposed	1	2	3
	34	18	52

Source: CDC, 1992, Table 4.

A subsequent, more complete report of the CDC-1 study published in the *New England Journal of Medicine* (Shands *et al*., 1980) contained an analysis of continual tampon use during the index menstrual period. These data, shown in Table A5.4 take into account the age-, sex-, and friend-matching used to recruit controls.

Question 3a Comment on the difference between Table A5.4 and Table A5.1. Which 2-by-2 table format is more appropriate for this study?
Question 3b How many cases used tampons continually?
Question 3c How many cases did not use tampons continually?
Question 3d How many controls used tampons continually?
Question 3e How many controls did not use tampons continually?
Question 3f Calculate the odds ratio and a 95% confidence interval for the odds ratio parameter. (Use formulas 13.21 and 13.22, respectively. You may use an epidemiologic calculator such as *EpiCalc2000* to check your calculations.)
Question 3g Calculate a *p* value for these data. Interpret your results. (Use formula 13.27 or an epidemiologic calculator for your computation.)
Question 3h Discuss your finding.
Question 3i In your opinion, have you been given enough information to decide whether the association is causal?

National publicity followed the 27 June report and continued almost daily throughout the summer of 1980. Though not addressed by the studies discussed so far, the lay press speculated that the then-new, highly absorbent tampons such as Rely brand might be responsible for TSS.

By 5 September 1980, CDC had received reports of 272 cases. At that time, CDC launched a second case–control study (CDC-2) to test the hypothesis that one or more brands of tampons might be more strongly associated with TSS than were other brands. The case group was 50 surviving females with onset of illness during July-August, 1980.

Question 4 List specific types of selection and information biases that may have been introduced by the *MMWR* reports, the intense publicity, and the nature of the disease?

For the moment, assume it is September 1980 and you have been asked to test the hypothesis that women of reproductive age (12–49 years) who use hypothetical

Brand X tampons during menses are at a greater risk of TSS than are women who use other brands.

Question 5 Assuming that you will conduct a case–control study using the 50 women with onset of TSS in July and August as your case group, who might you include in your control group? What are some of the possible sources of controls?

The (fictitious) marketing research data in Table A5.5 is intended to help you determine whether age is likely to be a confounder in a study of TSS and tampon brand. (No such marketing data on Rely was available to investigators in 1980.)

Table A5.5 indicates the distribution of women who used only one brand of tampon at the time of the study. Note that Brand X holds 14% of market share. However, it is much more popular among young women than among older women, as shown in Table A5.6.

Question 6a Calculate risk ratio for the data in Tables A5.5.
Question 6b Calculate risk ratio separately for the younger and older age groups in Table A5.6.
Question 6c Is there evidence of confounding?

Table A5.5 Distribution of hypothetical-brand-x-loyal tampon users and TSS cases, women age 12–49 years, united states, 1980.

Brand	Number of users	% of All users	Number of tss cases	Risk of tss per 100 000 users
X	4 900 000	14	452	9.2
All other	30 100 000	86	386	1.3
	35 000 000	100	838	

Table A5.6 Distribution of hypothetical-brand-x-loyal tampon users and TSS cases, women according to age.

Brand	Number of users	% of All users	Number of TSS cases	Risk of TSS per 100 000 users
12 to 29-Year-Old Women				
X	4 750 000	24	447	9.4
All other	15 250 000	76	287	1.9
	20 000 000	100	734	
30 to 49-Year-Old Women				
X	150 000	1	5	3.3
All other	14 850 000	99	99	0.67
	15 000 000	100	104	

Question 6d What can you do to control for this confounder?
Question 6e Why is age a confounder when considering the Brand-X-TSS association?
Question 7 What are the advantages and disadvantages of matching in case–control studies?
Question 8 In your study, would you match? Why or why not? On which characteristics would you match? What type of matching would you use?

For the CDC-2 study, cases were restricted to female TSS patients with onset of illness during July-August, 1980, who were reported to CDC by 5 September 1980, who survived their illness, and who met the CDC definition for TSS. Fifty cases met these eligibility criteria. For controls, the cases were asked to provide the names of three female friends or acquaintances of the same age (within 3 years) who lived in the same geographic area. The investigators used three controls rather than one for each case to increase their ability to detect an association between TSS and use of a particular brand of tampon, assuming that such an association existed.

Question 9 Do you agree with the CDC-2 investigators' decision to use friend controls?

One of the subanalyses of the CDC-2 study focused on the use of Rely brand tampons. This subanalysis excluded cases and controls who did not use tampons at all or who used more than one brand of tampon during the index menstrual period. As a result of these exclusions, some cases were matched to three controls and some were matched to only two controls. Data are shown in Table A5.7.

Use the following formula to calculate the odds ratios:

$$\text{Odds ratio} = \frac{\text{Number unexposed controls matched with exposed cases}}{\text{Number exposed controls matched with unexposed cases}}$$

Table A5.7 Exclusive use of rely brand tampons in matched cases and controls.

3 Controlspercase
NO. of controls using relybrand

	3 of 3	2 of 3	1 of 3	0 of 3
Case using rely	11		5	4
Case not using rely	01		1	1

2 Controlspercase
NO. of controls using relybrand

	2 of 2	1 of 2	0 of 2
Case using rely	3	3	7
Case not using rely	0	3	4

Example showing the calculation of the odds ratio for quadruplets (3:1 matched sets)

No. of unexposed controls matched with exposed cases = (0)(1)+(1)(1) + (2)(5) + (3)(4) = 23

No. of exposed controls matched with unexposed cases = (3)(0) + (2)(1) + (1)(1)+(0)(1) = 3

Odds ratio = 23/3 = 7.67

Question 10 Calculate the odds ratio for the 2:1 matched sets. Interpret your finding. Is it consistent with what was found for the quadruplets?

To properly analyze these data, we would have to combine the information from all the matched pairs while controlling for additional confounders. One way to accomplish this is to use each matched tuple as separate strata and then apply the Mantel-Haenszel methods discussed in Chapter 14. This is more or less what was done in the final analysis in an article published in *JAMA* by Schlech and co-workers (1982). The information from all pairs, triplets, and quadruplets was combined to yield a Mantel-Haenszel summary odds ratio of 7.7 (99% confidence interval: 2.1, 27.8, $p < 0.0001$).

The CDC investigators wished to test the hypothesis that one or more brands of tampons might be more strongly associated with TSS than other brands. To test that hypothesis, data were analyzed using a logistic regression model for matched data. The results of these analyses are shown in Table A5.8.

Question 11 Only the Rely brand tampons showed an increase in risk: All other tampon brands have odds ratios less than 1. Can manufacturers of tampons other than Rely claim that their brands protect against TSS?

Conclusion

The results of the CDC2 study were initially published in the *MMWR* of 19 September 1980. The study showed a strong and statistically significant association between

Table A5.8 Odds ratios for toxic shock syndrome among tampon users.

Brand	Odds ratio	(99% Confidence interval)
Rely	7.7	(2.1, 27.9)
Playtex	0.7	(0.2, 2.7)
Tampax	0.1	(0.02, 1.0)
Kotex	0.2	(0.01, 2.8)
OB	0.3	(0.002, 4.4)

Source: Schlech *et al*., 1982, Table 2.

Rely brand tampons and TSS. On 22 September 1980, after discussions involving the CDC, the U.S. Food and Drug Administration (FDA), and Proctor and Gamble (the manufacturers of Rely), the company voluntarily withdrew Rely tampons from the market.

At about the same time, CDC stopped accepting case reports of TSS and instead referred persons who wished to report a case to their state health departments. In addition, the number of menstruating women using tampons declined from about 70% to about 50%. Subsequently, the number of TSS cases reported to the CDC declined. While CDC attributed the decline to the withdrawal of Rely and overall reduction in tampon use, critics have charged that the decrease in reported cases may have been due to changes in the reporting system.

In 1982, the FDA required the labeling of tampons to advise women to use the lowest absorbency tampons compatible with their needs. On 1 January 1981, toxic shock syndrome became a nationally reportable disease. In March, 1990, ten years after the original epidemic of TSS, the FDA instituted standardized absorbency labeling of tampons (CDC, 1990; Schuchat and Broome, 1991).

References

CDC. (1980, May 23). Toxic shock syndrome—United States. *MMWR*, **29**, 297–299. (Initial CDC report; stimulated publicity and reports from around the country).

CDC. (1980, June, 27). Toxic shock syndrome-United States. *MMWR*, **29**, 229–230. (CDC-1).

CDC. (1990). Historical perspectives reduced incidence of menstrual toxic-shock syndrome—United States, 1980–1990. *MMWR*, **39**, 421–423. (Brief 10-year perspective.)

CDC. (1992). *Tampons and toxic shock syndrome*. Washington, DC: Association of Teachers of Preventive Medicine. (Training materials that served as foundation for current case study.)

Davis, J. P., Chesney, P. J., Wand, P.J., and LaVenture, M. (1980). Toxic-shock syndrome: Epidemiologic features, recurrence, risk factors, and prevention. *New England Journal of Medicine*, **303**, 1429–1435. (Wisconsin study.)

Harvey, M., Horwitz, R.I., and Feinstein, A. R. (1982). Toxic shock and tampons. Evaluation of the epidemiologic evidence. *JAMA*, **248**, 840–846. (Criticism of studies.)

Schuchat, A., and Broome, C.V. (1991). Toxic shock syndrome and tampons. *Epidemiologic Reviews*, **13**, 99–112. (Expanded 10-year perspective.)

Schlech, W.F., Shands, K. N., Reingold, A.L., Dan, B. B., *et al*. (1982). Risk factors for development of toxic shock syndrome: Association with a tampon brand. *JAMA*, **248**, 835–839. (General publication of results from CDC-2 plus additional laboratory studies and results by absorbency indices.)

Shands, K.N., Schmid, G.P., Dan, B.B., *et al*. (1980). Toxic-shock syndrome in menstruating women: Association with tampon use and *Staphylococcus aureus* and clinical features in 52 cases. *New England Journal of Medicine*, **303**, 1436–1442. (General publication of results from CDC-1.)

Todd, J., Fishaut, M., Kapral, F., and Welch, T. (1978). Toxic-shock syndrome associated with phage-group-I staphylococci. *Lancet*, **2**, 1116–1118. (Initial publication about toxic-shock syndrome.)

Index

Note: Page numbers in *italics* refer to Figures; those in **bold** to Tables.

absolute comparison, 75
absolute measures of effect, 75, 153, 168
acquired immunity, 261, 263–4, *264*
active immunization, 264
active surveillance, 105–6
actuarial adjustment, 61
actuarial method of survival analysis, 360–2
adjustment for multiple factors, 89–90
admissibility (eligibility) criteria, 149–50, 165
adoptive immunization, 265
Advisory Committee to the U.S. Surgeon General, 1964, Report of, 49–50
age adjustment (age standardization), 84
age intervals, 375
Age of Enlightenment, 10, 31
agents, 45, 256
agreement, 223, 224–8
aggregate-level data, 116
aggregate-level variables, 120–1
aggregation bias, 119–20
all-or-none phenomenon, 43
alternative hypothesis, 312
ambidirectional cohort studies, 164–5, *164*
analyzed-as-randomized analysis, 151–2
analytic epidemiology, 104–5
annual mortality rate, 375, 380
antigens, 264
antilogs, 307
apparent prevalence, 231–2
artifactual fluctuations in reported rates, 353–4
attention bias, 146
attributable fraction
  in exposed cases, 81
  in population, 79–81, **80**, **81**
average risk, 43, 69

B lymphocytes, 264
baseline comparisons, 149
Bayes factor, 320
Bayesian formula,
  for predictive value positive, 233
  for prevalence, 232–3
bell-shaped curve, 315
Berkson's bias, 211
Bernouilli's binomial distribution, 314
bias, 83
binomial test, 314
biological cycles, 261
biological interdependence, 337–8
biological variation, 205
birth certificates, 107
blinding (masking), 147, 151
blocked designs, 148
bloodletting, 15–16, *16*
Brennan-Prediger kappa coefficient, 228
British Doctors Study, 28–30, 163

carriers, 258
cart-before-the-horse bias, 134
case definitions, 151, 163–4, 182, 274, 347–51, 354
  AIDS (example), 350
  chronic fatique syndrome (example), 348–50, **349**
  classification of status based on certainty, 350–1
  multiple-choice criteria, 348
case-control odds ratio as relative risks, 193–4
case-control studies, 135–7, 180–200
  data analysis, 186–92
  identifying cases and controls, 182–5
  obtaining information on exposure, 185–6
case series, 23–4, 105, 135
cases, ascertainment of, 182–3
catchment area, 183
causal complements, 43
  prevalence of, 43–4
causal inference, 48–74
causal inference framework (Hill), 50–6
  analogy, 55–6
  biological gradient, 51–3
  biological plausibility, 53
  coherence, 53–5
  consistency, 50–1
  experimentation, 55
  specificity, 51

Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology, Third Edition.
B. Burt Gerstman.

© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

strength, 50
temporality, 51
causal interactions, 43, 337
causal mechanism, 127
causal models, 41–7
causal pies, 42
causal web, 44–5, *45*
cause, definition of, 41–2
cellular and physiologic barriers to infection, 263
chance, 205, 225, *226*
chemical barriers to infection, 263
chi-square teste
of association, 315–17
for statistical interaction, 340–1
cholera, 19–24, 110, 111, 114, 119, *120*, 276–7, 280
chronic carriers, 258
Clever Hans effect (obsequiousness bias), 213
clinical trials, 137, 142
closed population, 67, *67*, 101
clustering, 389–90, 394–5
Cochran-Mantel-Haenszel chi-square statistic, 368–9
Cochran–Mantel–Haenszel summary risk difference, 366–8
Cochran–Mantel–Haenszel summary risk difference estimate, 330
cofactors, 127, 128
cohort, definition, 67, *67*, 101, 159
cohort analysis, 22–3
cohort life table *see* generational life table
cohort studies, **52**, 135–7, 159, 180
common outcome, 102
common vehicle spread, 260–1
community trials, 137, 142
comparability, 148
component cause model, 42
confidence interval estimation, 304
confidence intervals, 206–8, *206*, 304–11, *304*
for odds ratios, independent samples, 310
for odds ratios, matched pairs, 310–11
for proportion differences (risk differences and prevalence differences), 308–9
for proportion ratios (risk ratios and prevalence ratios), 306–7
for proportions (incidence proportion and prevalence), 304–5
small sample methods, 305–6
for rate differences, 309–10
for rate ratios, 308
for rates, 306
confidence level, 207
confirmed cases, 350
confounded comparisons, 83
confounders, 213
properties, 213, *214–15*
confounding, 165–6, 210, 213–16, 343–4
prevention of, 323–5
contact, 260
contagion variables, 121
contextual variables, 120–1
contributing component causes, 43
control group, 130–1, 146
control series, 135
control variables, 128
controls per case, numer of, 184
controls, selection of, 183
convalescent carriers, 258
cooperation of study population, 163
counterfactual ideal, 165
Cox proportional hazards regression, 152
cross-over design, 147
cross-product ratio, 182
cross-sectional life tables, 373, 374
cross-sectional observation, 133–5
cross-sectional studies, 137
crude measures of association, 327
crude rate, 83
cumulative incidence proportion (risk) at time *t*, 365
cumulative incidence proportion (risk) difference, 365
cumulative incidence proportion (risk) ratio, 365
cumulative incidence sampling, 194
cumulative incidences, 69
current (cross-sectional) life tables, 373, 374
cutoff point for test results, 235–8
cyclozoonoses, 259

data, 128–9
Data and Safety Monitoring Board (DSMB), 130
data collection forms, 129
death certificates, 107–8
death rate, 375
decisions, types of, 48–9
definition of epidemiology, 2
degenerative diseases, 6
demographic changes, 20th century, 5–7
US, *6*
demographic transition, 5
denominator, 67, 101
dependent variable, 128
descriptive epidemiology, 104–25
definition, 104–5
ecological correlations, 116–19
epidemiologic variables, 108–15
detection bias, 134
diagnostic bias, 134
diagnostic suspicion bias, 213, 274
diagnostic tests, 223
dichotomous exposure, 186
differential misclassification, 212
dimensionality of ratio, 101
direct adjustment, 84–5
direct causes, 44
direct notification, 272
direct zoonoses, 259
disease, definition, 4, 74
disease patterns, 20th century changes in, 5–7
diseases of civilization, 6
Doll, Richard, 28–30, 163
dot maps, 277
double blinding, 151
Durkheim, Emile, 25

ecological analysis, 21, *22*
ecological correlations, 116–19
ecological fallacy (aggregation bias), 119–20
ecological studies, 116, 137

ecology of disease, 46
effect measure heterogeneity, 338, 340
effectiveness analysis, 151–2
empirical induction period, 39, *39*
endemic disease, definition, 19
endemic pattern, 113, *113*
environmental factors, 45
epidemic, definition, 2, 19
epidemic curve, 113, *113*, 114, *114*, 275, *276*, *277*
epidemiologic computing, 389
epidemiologic homeostasis, 46, *47*
epidemiologic iceberg, 40–1, *41*
epidemiologic studies, common types of, 137, *137*
epidemiologic surveillance systems, 272
epidemiologic transition, 5, 30
epidemiologic variables, 108–15, 275
equipoise, 130
estimates, 202–3, 303
estimation, 206–8, 304
ethical conduct (human subjects), 129–30, **129**
etiologic research, 126–9
exact binomial confidence limits, 305
exercises in measurement, 201
expected proportion of agreement due to chance, 227
expected survival time, 358
experimental error, 205
experimental studies (trials), 131–2, 142–57
  control group, 146–8
  data analysis, 152–5
  follow-up and oucome ascertainment, 151
  historical perspective, 144–6
  randomization and comparability, 148–9
  recruitment and eligibility criteria, 149–50
exposure, 74
  obtaining information on, 185–6
exposure odds ratio, 181
extraneous variables, 128

factorial designs, 147
false negatives (FN), 229
false positives (FP), 229
Farr, William, 17–19, *18*, 31, 119, *119*, *120*, 162
field trials, 137, 142, 143
Fisher's exact test, 317–18
fixed-level hypothesis (significance) testing, 207, 312
Fracastoro, Girolamo, 13
Framingham, 303
Framingham Heart Study, 160, 163, 167, 168, 169–70
frequency matching, 324
Frost, Wade Hampton, 162, 170–4, *173*

G index, 228
generational (cohort) life table, 373
  versus current (cross-sectional) life tables, 373
germ theory, 12–13
Goldberger, Joseph, 26–7, *27*, *162*
gradient of infection, 40
Graunt, John, 11–12
group comparability, checking, 149
Gwet's AC1, 228

Hawthorne effect, 146
health surveys, 108
health, definition, 3
healthy worker effect, 211
Henle, Jakob, 13
herd immunity, 265–6, *266*
  acquired, 265
  definition, 265
  epidemic modeling, 267
  innate, 265
  stemming an outbreak, 265–6
Hill, Austin Bradford, 28, 163
Hippocrates, 9, **10**, 161
history, 8–30, 31
  of experimental studies, 144–6
  observational cohort studies, 161
HIV/AIDS, 37–8, *39*
  case definitions, 350
  sexual transmission, 46, *46*
hospital admission bias, 211
host factors, 45
host immunity, 261–4
  acquired, 261, 263–4, *264*
  innate, 261, 262, *264*
  types, 261
hypothesis, definition, 56, 279
hypothesis, development of, 126–7, 279–80
hypothesis generation, 279, **279**
hypothesis statement, causal, 126–7
hypothesis testing, of statistical null hypotheses, 206, 208–9, 304, 312, 318–19
hypothetical populations, 303

illness, definition, 4
immunization, 264–5
immunocytes, 263
inapparent carriers, 258
incidence
  mathematical relationship between prevalence and, 103
  vs prevalence, 74, **74**
incidence density, 70–2
incidence density sampling, 193–4, *193*
incidence-prevalence bias, 134–5
incidence proportion (risk), 69, 102
incidence rate, 70–2, 102
incident cases, 182
incubation period, 36, 37
incubatory carriers, 285
independent rates, testing, 318–19
independent variable, 128
index group, 130
indirect adjustment, 85–9
indirect causes, 44
individual-level causes, 44
individual matching, 324
individual risk, 43
induction, 57
induction period, 36, 37, **38**, 39, *39*, 127
infections, 256
infectious agents, 256–9, **257**
  natural history within host, 261

infectious cycles, 261
infectious disease, 256, **257**
infectious disease process, 255–65
infectivity, 47
inflow, 73
information bias, 205, 210, 212–13
informed consent, 129
innate immunity, 261, 262, *264*
institutional review boards (IRBs), 129
integral group properties, 120
intention-to-treat (analyzed-as-randomized, effectiveness) analysis, 151–2
interaction, types of, 337–40
International Classification of Disease (ICD), 351–3, 354
interval estimation, 206, 304

John Henry effect, 146

Kaplan-Meier method of survival analysis, 362–4
Kaplan-Meier product-limit method, 362
kappa paradox, 227–8
kappa statistic, 225–7
killed vaccines, 264
Koch, Robert, 13

laboratory and environmental studies, 280–1
laboratory investigation, 287
Lane-Claypon, Janet Elizabeth, 162
latent period, 39, *39*
law of large numbers, 324
life expectancy, 7–8
life table, 373
  abridged, 375, 380–3
  complete, 374–5
    construction, 377–80
  stationary population, 374
lifestyle diseases, 31
Lind, James, 144–5
London Epidemiological Society, 2, 17, 31
longitudinal observations, 133–5
longitudinal studies, 137
Louis, Pierre Charles Alexandre, 14, 15–16, *15*
lower confidence limit (LCL), 206, *206*, 304
Lucretius, 12–13
Louis, Pierre Charles Alexandre, 162

macro-level causes, 44
Mantel–Haenszel adjusted rate difference parameter, 333
Mantel–Haenszel chi-square statistic, 316, 333–5, **334**
Mantel-Haenszel confidence interval for risk ratio, 328–9, **334**
Mantel-Haenszel methods, 323–36
Mantel-Haenszel risk ratio, 325–9
  confidence interval for, 328–9, **334**
  heterogeneity assumption, 326
  homogeneity assumption, 326–7
  mixing of effectx, 325–6
Mantel–Haenszel summary incidence proportion (risk) ratio, 327–8
Mantel–Haenszel summary odds ratio, 330–1
Mantel–Haenszel summary rate ratio, 331–3
Mantel–Haenszel test statistic, 329
mapping, 111
margin of error, 206
masking *see* blinding
mass action principle, 267
match tuples, 191–2
matched pairs, 189–91
matching, 165–6, 324–5
maternally derived passive immunity, 264
maximum-likelihood estimate (MLE) technique, 340–1
McNemar's test for matched pairs, 319
measure of association heterogeneity, 338
measures of association, 74–9
measures of disease frequency, 67–74
measures of effect, 74, 152–3, 166–170, 186–192
  absolute, 75–6
  relative, 76
measures of occurrence, 67
measures of potential impact, 79–82
Médecine d'observation, 14–16
medical ecology, 13
MedWatch (FDA), 107
metaphor in random and systematic error, 201–2, *202*, *203*
metazoonoses, 259
Méthode Numerique, La, 14–16
miasma theory of transmission, 20
micro-level causes, 44
Minimum Bayes factors, 319–21
  calculation, 320–1
misclassification bias, 212
misinterpretation, 209
modified life table, 360
modified live vaccines, 264
morbidity, evolving patterns of, 5–7, 30–31
mortality, 5–7
  evolving patterns of, 30–1
  leading causes, US, **5**
  trends since, 1950, 7, *8*
multilevel analysis, 121
multiple levels of exposure, 186–7

national morbidity and mortality databases, 107–8
National Notifiable Disease Surveillance System, 107
natural experiment, 145–6
natural history of disease, 36–40
natural logarithm, 306
nested case-control studies, 183
nondifferential misclassification, 151, 212–13, *212*
non-experimental epidemiologic studies, 131
nonrandomized trials, 131, 144
nonresponse bias, 211
normal approximation, 305, 314
nosocomial infections, 259
notation, 181, 203
null distribution, 312, *313*, 314
null hypothesis, 208, 209, 312, 315, 317
numerator, 67, 101

obsequiousness bias, 213
observation, 56

observational cohort studies, 159–79
  assembling and following a cohort, 163–4
  birth cohort study (Frost), 170–4
  confounding, 165–6
  data analysis, 166–70
  historical perspective, 161–3
  prospective, retrospective and ambidirectional cohorts, 164–5
observational studies, 131–2
observed proportion of agreement, 226
observed rate, 357–8
odds, 102
odds ratio, 77–8, *78*, 181
odds ratio estimate, 310
odds ratio estimator, 310
open population, 67, *68*, 72
OpenEpi.com, 153–4, *154*, **154**, 155, *167*, *168*, *169*, 305
outbreaks
  drug-disease outbreak (case study), 283–6
  food borne outbreak (case study), 286–291
  goals and methods of investigations, 272–3
  initial detection, 272
  investigatory steps, 273–82
    communication of findings, 282
    control and prevention methods, 281
    descriptive epidemiologic studies, 275–9
    establishment of outbreak, 273–4
    hypothesis development, 279–80
    hypothesis evaluation, 280–1
    preparation for field work, 273
    verification of diagnoses, 274,
outflow, 73
overmatching, 324
p-values, 208–9, 312–19
  basis of, 312–13
  fallacies of, 313

pandemic, definition, 19
parallel design, 147
parameter, 202–3, 303
Paré, Ambroise, 145
passive immunization, 264
passive surveillance, 105, 107
Pasteur, Louis, 13
pathogenicity, 47
Pearson's uncorrected chi-square statistic, 316
per-protocol analysis (efficacy analysis), 151–2
period prevalence, 73
person variables, 108, **109**, 109–10, 277–9
person-time unit, 70
physical barriers to infection, 262
Pinel, Philippe, 14, 162
place variables, 108, **110**, 111, 277
placebo effect, 146
point epidemic pattern, 113, *113*
point estimate, 206
point estimation, 206, 304
point estimator, 309
point prevalence *see* prevalence
point-source epidemics, 277
Poisson distribution, 315, 386–9, *388*, 395
  fitting, 390
  goodness of fit, 390–4
Poisson formula, 386
  calculating expected number of cases, 387–8
  use of, 387
polls, political, 210–11
Popper, Karl, 57–8
population, 303
portals of entry and exit, 259–60
possible cases, 350
posterior odds, 320
potential confounders, 128
Pott, Percival, 10–11, 162
precision of estimate, 207
preclinical phase, 36
predicted probability of death, 376–7, 382
  in first year of life, 376
prediction, 56
predictive value of a negative test (PVNT), 230–1
  vs prevalence, 234–5, *235*
predictive value positive of a positive test (PVPT), 230–1
  vs prevalence, 232–4, *234*
prevalence, 72–4, *74*
  Bayesian formulas for, 232–3
  vs incidence, 74, **74**
  mathematical relationship between incidence and, 103
  vs predictive value of a negative test (PVNT), 234–5, *235*
  vs predictive value positive of a positive test (PVPT), 232–4, *234*
prevalence difference, 75
prevalence-incidence bias, 134–5, 211
prevalence proportion *see* prevalence
prevalence ratio, 76–7
prevalent cases, 183
preventable fraction, 82
  in the population, 82
  in the unexposed, 82
primary prevention, 40
prior odds, 320
probability, 204–5
probabilities of death, 375
probable cases, 350
propagating epidemic pattern, 113, *113*, 277
proportion, 101–2
  of specific negative agreement, 228
  of specific positive agreement, 228
  testing, 314–15
proportion ratio parameter, 306
proportional-hazards regression, 166
prospective cohort studies, 164–5, *164*
public health, definition, 2–3
publicity bias, 211, 274

quadratic method, 305
qualitative method, 280
quantitative epidemiologic investigations, 280

Ramazzini, Bernardino, 10, 161–2
random distribution in time and space, 385–6
random error (imprecision), 201–3, 204–9

random noise, 205
randomization, 148, 324
randomized controlled trial (RCT), 144, 148
randomized trials, 131
rare outcome, 102
rate, 101
  constant, 102–3
rate adjustment, 82–4
rate difference, 75, 76, 153, 168, 309–10, 318–9
rate ratio, 75, 76, 77, 152, 166–8, 308, 318–9
rates by region, 277
rates, predicting probabilities from, 375–6
ratios, 67, 101
real population, 303
recall bias, 213
recruitment, 149–51
rectangularization, 374
Reed-Frost model, 267
reference (control) group, 130–1
reference population, 84
referrent (control) groups, 146
refutation, 57–8
regression models, 325
relative comparison, 75
relative measures of effect, 152
relative risks (RR), 76
vs relative risk difference, 77
reliability (agreement), 223, 224–8
reproducibility, 224
reservoirs, 257
resolution, 36
restriction, 324
retrospective cohort studies (historical cohort studies), 164–5, *164*
reverse-causality bias, 134
risk, perception of, 102
risk difference, 75, 76, 153, 168–9, 308–9, 315
  vs risk ratio, 78–9
  vs relative risk difference (RRD), 77
  at selected points in time, 381–2
risk multiplier, 76
risk ratio, 76, 77, 152, 166, 306–7, 315
  vs relative risk difference (RRD), 77
  vs risk difference, 78–9
  at selected points in time, 381–2
Russell, Bertrand, 57

Salmon, Daniel E., 13
sample size, 127
  considerations, 184–5
  requirements, 155
saprozoonoses, 259
screening, 222–54
  HIV/AIDS case study, 244–8
secondary prevention, 40
selection bias, 163, 205
sensitivity (SEN), 229, 230
serial transfer, 261
shoe-leather epidemiology, 274
'shooting for', 203
sickness, definition, 4
significance testing, 208, 312
signs, 223
Simpson's paradox, 325, **326**
smoking, 28–9, *29*, **29**, *30*
Snow, John, 13, 19–24, *20*, 31, 111, *112*, 276, 277
  publication, 24
  cohort analysis, 22–3
  case series, 23–4
  ecological analysis, 21, *22*
source population, 149, 183
specificity (SPEC), 229, 230
spectrum of disease, 40
sporadic occurrence, 113, *113*
stage of clinical disease, 36
stage of subclinical disease, 36
stage of susceptibility, 36
stages of disease, 36–9, *37*
stages of prevention, 40
standard error, 308, 330
standard error of the natural log of Mantel–Haenszel incidence proportion ratio, 328
standard error of the natural log of the odds ratio, 310
standard error of the natural log of the proportion ratio, 306
standard error of the proportion, 305, 314
standard error of the rate difference, 309
standard error of the risk difference, 309
standard million, 84
standardized morbidity ratio (SMR), 85–6
stationary population, 67
statistical adjustments, 166
statistical inference, 153, 205–6, 303–4
statistical interaction, 337, 338–40
statistical testing, fallacies of, 313
strategy for stratified analysis, 342–4
stratification, 325
stratified analysis, strategy for, 342–4
stratified designs, 148
stratifying rates by follow-up time, 359–60
stratum, 83
study population, 84
sufficient causal mechanism, 42, *42*
sulfur oxide air pollution, 47
surveillance systems, 105–7
Surveillance, Epidemiology and End Results (SEER), 106–7
survival analysis, 356–9
  actuarial method, 360–2
  comparison to two groups, 364
  Kaplan-Meier method, 362–4
survival function, 362
  comparison, 366–8
Sydenham, Thomas, 10
symptomatic cases, 258
symptoms, 223
synergism, 337
systematic error (bias), 201–3, 209–11

T lymphocytes, 264
tertiary prevention, 40
therapeutic derived immunity, 264
time variables, 108, 111–15, **113**, 275–7
toxoids, 264

transmission, 260–1
  dynamics, 260–1
  mode of, 260
trials, 142
triple blinding, 151
true negatives (TN), 229
true positives (TP), 228
true prevalence, 231–2

unit of concern, 2
unit of observation, 116, 132–3
upper confidence limit (UCL), 206, *206*, 304
US life table survival curves, 373, **374**
uses of epidemiology, 4–5, **4**

vaccines, 264
validity, 223, 228–38
van Helmont, 144
variables, 128
vectors, 260
vehicles, 260
verification, 56
virulence, 47

Weinberg, Wilhelm, 162
WinPEPI, 153, 155, *155*, 170, 228, 305, 315, 364, 389
withdrawal bias, 211
withdrawals, 163

Yates's continuity-corrected chi-square statistic, 316
yellow shank disease, chicken, 44

z statistic, 318
zoonoses, 258–9