Fundamentals of Epidemiology and Biostatistics: Combining the Basics

Ray M. Merrill, PhD, MPH, MS
Brigham Young University
Provo, Utah

JONES & BARTLETT
LEARNING

47728_FMXX_Merrill.indd 1 4/19/12 3:05 PM

World Headquarters
Jones & Bartlett Learning
5 Wall Street
Burlington, MA 01803
978-443-5000
info@jblearning.com
www.jblearning.com

Jones & Bartlett Learning books and products are available through most bookstores and online booksellers. To contact Jones & Bartlett Learning directly, call 800-832-0034, fax 978-443-8000, or visit our website, www.jblearning.com.

Substantial discounts on bulk quantities of Jones & Bartlett Learning publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones & Bartlett Learning via the above contact information or send an email to specialsales@jblearning.com.

Copyright © 2013 by Jones & Bartlett Learning, LLC, an Ascend Learning Company

All rights reserved. No part of the material protected by this copyright may be reproduced or utilized in any form, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owner.

This publication is designed to provide accurate and authoritative information in regard to the Subject Matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the service of a competent professional person should be sought.

Fundamentals of Epidemiology and Biostatistics: Combining the Basics is an independent publication and has not been authorized, sponsored, or otherwise approved by the owners of the trademarks or service marks referenced in this product.

Some images in this book feature models. These models do not necessarily endorse, represent, or participate in the activities represented in the images.

Production Credits
Publisher: Michael Brown
Managing Editor: Maro Gartside
Editorial Assistant: Chloe Falivene
Editorial Assistant: Kayla Dos Santos
Associate Production Editor: Rebekah Linga
Senior Marketing Manager: Sophie Fleck Teague
Manufacturing and Inventory Control Supervisor: Amy Bacus
Composition: Circle Graphics, Inc.
Cover Design: Michael O'Donnell
Cover Image: © Terry Chan/ShutterStock, Inc.
Printing and Binding: Edwards Brothers Malloy
Cover Printing: Edwards Brothers Malloy

To order this product, use ISBN: 978-1-4496-6753-5

Library of Congress Cataloging-in-Publication Data
Merrill, Ray M.
Fundamentals of epidemiology and biostatistics : combining the basics / Ray M. Merrill.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-4496-4772-8 (pbk.) — ISBN 1-4496-4772-3 (pbk.)
I. Title.
[DNLM: 1. Epidemiologic Methods. WA 950]
614.4—dc23
2011047992

6048

Printed in the United States of America
16 15 14 13 12 10 9 8 7 6 5 4 3 2 1

DEDICATION

To Amy Valette

47728_FMXX_Merrill.indd 3 4/19/12 3:05 PM

CONTENTS

Preface xi
About the Author xv

CHAPTER 1 Introduction 1

Health 1
Epidemiology 4
Transmission of Disease 9
Descriptive Epidemiology 10
Analytic Epidemiology 13
Study Designs 14
Scientific Method 15
Biostatistics 16
Evaluating the Literature 17
Exercises 19
References 20

CHAPTER 2 Data and Descriptive Measures 25

Data and Related Concepts 25
Populations, Samples, and Random Sampling 27
Scales of Measurement 28
Summarizing and Presenting Nominal and Ordinal Data 28
Summarizing and Presenting Numerical Data 41
Exercises 50
References 54

CHAPTER 3 Standardizing Rates 57

Age-Specific Rates 58
Direct Method of Standardization 58
Indirect Method of Standardization 61
Uses of Standardized Rates 62
Exercises 66
References 67

47728_FMXX_Merrill.indd 5 4/19/12 3:05 PM

CHAPTER 4 Describing Data with Graphs 69
Arithmetic-Scale and Logarithmic-Scale Line Graphs 71
Bar Chart 73
Pie Chart 75
Histogram 76
Frequency Polygon 79
Area Map 80
Spot Map 81
Stem-and-Leaf Plot 82
Box Plot 85
Scatter Plot 87
Population Pyramid 87
Exercises 90
References 95

CHAPTER 5 Life Tables* 97
Calculation of the Probability of Dying (q_x) 98
Calculation of the Remaining Life Table 98
Abridging the Complete Life Table 100
Multiple Cause Life Table 104
Years of Potential Life Lost 108
Exercises 115
References 117

CHAPTER 6 Probability 119
Probability Concepts 119
Mutually Exclusive Events versus Independent Events 122
Conditional Probabilities 122
Rule for Multiplying Probabilities 123
Independent Events 123
Bayes' Theorem 126
Clinical Epidemiology 129
Methods of Sampling 130
Exercises 133
References 135

CHAPTER 7 Random Variables and Probability Distributions 137
Random Variables 137
Probability Distributions for Discrete Random Variables 138

The Binomial Probability Distribution 139
The Poisson Probability Distribution 141
Probability Distributions for Continuous Random Variables 142
Normal Probability Distribution 142
Normal Approximation to the Binomial 144
Sampling Distribution 148
Sampling Distribution of the Mean 148
The Central Limit Theorem 149
Exercises 151
References 154

CHAPTER 8 Estimation and Hypothesis Testing 155

Estimators 155
Confidence Intervals for Normal Means and Known Variance 155
t Distribution 156
Confidence Intervals for Normal Means and Unknown Variance 157
Confidence Intervals for the Binomial with Large Sample Sizes 158
Hypothesis Testing 159
The *P* Value 160
Type I and Type II Errors 160
Research Question about the Mean in One Group 161
Research Question about a Proportion in One Group 164
Research Question about Means in Two Separate Groups 169
Research Question about Proportions in Two Independent Groups 172
Evaluating Differences Using Error Bar Graphs 174
Chi-Square Distribution 176
Research Question about Means in Three or More Groups 178
Exercises 182
References 185

CHAPTER 9 Sample Size 187

Sample Size Techniques for Analytic Studies 187
Loss to Follow-Up 191
Fixed Sample Size 192
Ways to Minimize Sample Size and Maximize Power 193
Sample Size Techniques for Descriptive Studies 194

Exercises 198
References 199

CHAPTER 10 Study Designs 201

Study Design 201
Cross-Sectional Study Design 204
Ecologic Study Design 204
Case-Control Study Design 205
Cohort Study Design 207
Nested Case-Control Study Design 209
Chance, Bias, and Confounding 209
Chance 209
Bias 211
Confounding 214
Exercises 215
References 217

CHAPTER 11 Statistical Measures of Association Among Variables 219

Classification of Statistical Techniques and Tests by Variable Type 219
Measures of Association for Two-Level Nominal Exposure and Outcome Variables 233
Case-Control Study 234
Cohort Study 234
Cross-Sectional Study 235
A Note on the Risk and Odds Ratios 244
Some Cohort-Based Measures 244
Sample Size Estimation for the Correlation Coefficient 248
Exercises 250
References 256

CHAPTER 12 Experimental Studies 259

Experimental Study Designs 259
Random Assignment 260
Nonrandom Assignment 262
Blinding 263
Designing a Clinical Trial 265
Other Selected Types of Experimental Studies 269

Exercises 272
References 273

CHAPTER 13 **Survival Analysis*** **275**

The Life Table Method 276
The Product Limit Method 283
Log-Rank Test 287
Exercises 292
References 296

CHAPTER 14 **Cause and Effect** **297**

Causal Inference 297
Causal Models 302
Modes of Disease Transmission 303
Risk Factors 305
Web of Causation 306
Exercises 310
References 312

Appendix A: SAS Procedure Code 313

Appendix B: Tables 325

Appendix C: Model Building 353

Appendix D: Answers to Chapter Exercises 357

Index 411

*Chapters 5 and 13 are more statistically rigorous and may be considered optional in a beginning class. Skipping these chapters does not interfere with the flow of the book.

PREFACE

Fundamentals of Epidemiology and Biostatistics: Combining the Basics is an introductory text for students in epidemiology, clinical medicine, health policy and management, nursing, and various allied health fields. Students and professionals in all these settings should have knowledge of the basic principles of epidemiology and biostatistics. There are many subfields of public health and medical research, but epidemiology and biostatistics are fundamental to all these areas of study.

This book is intended to be used in a single course offered during a block or semester. Its primary purpose is to present the distinct and complementary roles of epidemiology and biostatistics to public health and clinical medicine in the context of substantive, interesting, and current applications. *Fundamentals of Epidemiology and Biostatistics: Combining the Basics* presents the material in a clear, concise, and accurate manner. Readers of this book will gain an appreciation of the fact that identifying, understanding, and solving public and clinical health problems often require thinking in terms of principles and practices in the areas of epidemiology and biostatistics.

Some of the special features of this book include the following:

Topics: This text includes fundamental concepts and methods that are required for a basic understanding and interpretation of epidemiology and biostatistics for public health and clinical medicine. The outline follows a logical sequence.

Substantive examples: Practical and relevant public health and clinical examples are used in each chapter.

Actual data: Many of the examples presented are based on actual data involving health-related states or events.

Assignment problems: Each chapter has a set of questions to determine students' mastery of the material.

Graphical excellence: Introductory statistical graphics that can be produced with spreadsheet or charting software are presented in order to effectively present data.

Computer applications: Computer applications using SAS are used to demonstrate data analysis throughout the book; procedure code is provided in Appendix A.

Statistical inference: Concepts of sampling, hypothesis testing, *P* values, and confidence intervals are presented.

Causal inference: Ways to support conclusions about cause-and-effect relationships are presented.

47728_FMXX_Merrill.indd 11 4/19/12 3:05 PM

Structure: The book sets off definitions, historical information, and certain formulas in boxes, and key words are in italic type in order to enhance readability and strengthen prominent points.

This book has four major biostatistics themes: descriptive statistical procedures (Chapters 2–5), probability (Chapters 6 and 7), statistical inference (Chapter 8), and statistical techniques for determining appropriate sample size and assessing data (Chapters 9, 11, and 13).

In the chapters covering descriptive statistics, we define terms important in the application of epidemiology and biostatistics, present scales of measurement and ways to summarize and present nominal and ordinal data, describe two methods to standardize rates, introduce life tables, present the method for calculating years of potential life lost, and describe standard ways to summarize and present numerical data. We also introduce statistical notation and present several types of graphs that are useful for describing the distribution of health-related states and events.

In the two chapters covering probability, we introduce probability concepts, probability distributions, probability distribution for sample statistics (called sampling distributions), Bayes' Theorem, and selected sampling methods. Statistical inference is introduced in a single chapter that addresses methods for estimating population parameters and how to examine and test hypotheses. Finally, the statistical techniques part of the book includes various methods for calculating sample size for different types of research questions and selected approaches for analyzing research questions about relationships among variables and about patient survival across time.

These four major biostatistics themes play a fundamental role in conducting observational and experimental epidemiologic studies, which investigate the distribution and determinants of health-related states or events in human populations. The primary purpose of this study is to prevent and control health problems. The study design is central to epidemiology. Several study designs employed in epidemiology are introduced and described, along with the quantitative elements of these designs. Chapter 1 provides a description of the distinct yet complementary roles of epidemiology and biostatistics in public health and clinical medicine and the basic study designs used in epidemiologic research. Chapter 10 focuses on selected observational study designs, while Chapter 12 focuses on experimental study designs. Finally, because drawing conclusions about causality is critical in epidemiology for developing appropriate disease and health-related prevention and control measures, this topic concludes the book (Chapter 14). Criteria for establishing cause-and-effect relationships are discussed, and selected causal models are presented.

47728_FMXX_Merrill.indd 12 4/19/12 3:05 PM

After completing this book, the student should be able to:

- Describe the distinct but complementary roles of epidemiology and biostatistics in public health and clinical medicine.
- Apply descriptive epidemiology and biostatistics procedures to understand and describe the distribution of health-related states or events.
- Apply probability concepts and probability distribution theory to understanding and describing health problems, identifying the efficacy of medical screening and testing procedures, and sampling in order to obtain a representative group of study participants.
- Derive estimators from a sample that is used to estimate an unknown population parameter and measure the precision in the estimators using confidence intervals.
- Apply the steps of hypothesis testing to answer research questions about means and proportions in single groups, in two independent groups in one group, in two dependent groups, in more than two independent groups, or in more than two dependent groups.
- Determine appropriate sample size and assess various types of research questions.
- Identify and apply appropriate observational or experimental study designs to answer specific public health or medical research questions.
- Estimate survival curves using the life table (actuarial) method or the product-limit (Kaplan-Meier) method and evaluate statistical significance of differences between curves using the log-rank test.
- Apply causal theory and causal criteria to identify determinants of health-related states or events.
- Apply the study of the distribution and determinants of health-related states or events to prevent and control health problems.
- Apply SAS in studying the distributions and determinants of health-related states or events in human populations.

ABOUT THE AUTHOR

Ray M. Merrill, PhD, MPH, MS, is a professor of epidemiology and biostatistics at Brigham Young University. He is also an adjunct professor in the Department of Family and Preventive Medicine and the Department of Health Promotion and Education at the University of Utah. He is a former cancer prevention fellow at the U.S. National Cancer Institute and a visiting scientist in the Unit of Epidemiology for Cancer Prevention at the International Agency for Research on Cancer, Lyon, France. He has won several awards for his research in epidemiology and is currently a fellow of the American College of Epidemiology and of the American Academy of Health Behavior. He teaches classes in epidemiology and biostatistics and is the author of over 200 peer-reviewed publications and 5 books.

47728_FMXX_Merrill.indd 15 4/19/12 3:05 PM