
EPIDEMIOLOGY

Beyond the Basics

THIRD EDITION

MoySES Szklo, MD, DrPH

Professor of Epidemiology and Medicine

The Johns Hopkins University

Editor-in-Chief, *American Journal of Epidemiology*

Baltimore, MD

F. Javier Nieto, MD, PhD

Helphaer Professor of Public Health

Professor and Chair

Department of Population Health Sciences

University of Wisconsin School of Medicine and Public Health

Madison, WI



JONES & BARTLETT
LEARNING

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 © 2014 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.

Epidemiology: Beyond the Basics, Third Edition 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.

Production Credits

Chief Executive Officer: Ty Field
President: James Homer
SVP, Editor-in-Chief: Michael Johnson
SVP, Chief Marketing Officer: Alison M. Pendergast
Publisher: Michael Brown
Associate Editor: Teresa Reilly
Editorial Assistant: Chloe Falivene
Editorial Assistant: Kayla Dos Santos
Director of Production: Amy Rose
Production Manager: Tracey McCrea
Production Assistant: Eileen Worthley
Senior Marketing Manager: Sophie Fleck Teague
Manufacturing and Inventory Control Supervisor: Amy Bacus
Composition: Lapiz, Inc.
Cover Design: Kristin E. Parker
Cover Image: © BioMedical/Shutterstock, Inc.
Printing and Binding: Edwards Brothers Malloy
Cover Printing: Edwards Brothers Malloy

Library of Congress Cataloging-in-Publication Data

Szklo, M. (Moyses)
Epidemiology : beyond the basics / Moyses Szklo and F. Javier Nieto.—3rd ed.
p. : cm.
Includes bibliographical references and index.
ISBN 978-1-4496-0469-1 (pbk.) — ISBN 1-4496-0469-2 (pbk.)
I. Nieto, F. Javier. II. Title.
[DNLM: 1. Epidemiologic Methods. 2. Epidemiology. WA 950]
614.4—dc23

2012015799

6048

Printed in the United States of America

16 15 14 13 12 10 9 8 7 6 5 4 3 2 1

To Hilda and Marion

Contents

Preface	ix
Acknowledgments	xii
About the Authors	xiii

PART ONE Introduction 1

CHAPTER 1 Basic Study Designs in Analytical Epidemiology 3

1.1	Introduction: Descriptive and Analytical Epidemiology	3
1.2	Analysis of Age, Birth Cohort, and Period Effects	4
1.3	Ecologic Studies	14
1.4	Studies Based on Individuals as Observation Units	19
	References	39
	Exercises	42

PART TWO Measures of Disease Occurrence and Association 45

CHAPTER 2 Measuring Disease Occurrence 47

2.1	Introduction	47
2.2	Measures of Incidence	49
2.3	Measures of Prevalence	71
2.4	Odds	73
	References	74
	Exercises	75

CHAPTER 3 Measuring Associations Between Exposures and Outcomes 79

3.1	Introduction	79
3.2	Measuring Associations in a Cohort Study	79
3.3	Cross-Sectional Studies: Point Prevalence Rate Ratio	90
3.4	Measuring Associations in Case-Control Studies	90
3.5	Assessing The Strength of Associations	101
	References	104
	Exercises	105

PART THREE Threats to Validity and Issues of Interpretation 107**CHAPTER 4 Understanding Lack of Validity: Bias 109**

- 4.1 Overview 109
- 4.2 Selection Bias 111
- 4.3 Information Bias 116
- 4.4 Combined Selection/Information Biases 133
- References 147
- Exercises 150

CHAPTER 5 Identifying Noncausal Associations: Confounding 153

- 5.1 Introduction 153
- 5.2 The Nature of the Association Between
the Confounder, the Exposure, and the Outcome 156
- 5.3 Theoretical and Graphical Aids to Frame Confounding 162
- 5.4 Assessing the Presence of Confounding 164
- 5.5 Additional Issues Related to Confounding 171
- 5.6 Conclusion 179
- References 180
- Exercises 182

CHAPTER 6 Defining and Assessing Heterogeneity of Effects: Interaction 185

- 6.1 Introduction 185
- 6.2 Defining and Measuring Effect 186
- 6.3 Strategies to Evaluate Interaction 186
- 6.4 Assessment of Interaction in Case-Control Studies 196
- 6.5 More on the Interchangeability of the Definitions of Interaction 203
- 6.6 Which is the Relevant Model? Additive Versus Multiplicative Interaction 205
- 6.7 The Nature and Reciprocity of Interaction 207
- 6.8 Interaction, Confounding Effect, and Adjustment 211
- 6.9 Statistical Modeling and Statistical Tests for Interaction 212
- 6.10 Interpreting Interaction 214
- 6.11 Interaction and Search for New Risk Factors in Low-Risk Groups 219
- 6.12 Interaction and "Representativeness" of Associations 221
- References 222
- Exercises 224

PART FOUR Dealing with Threats to Validity 227**CHAPTER 7 Stratification and Adjustment: Multivariate Analysis in Epidemiology 229**

- 7.1 Introduction 229

7.2	Stratification and Adjustment Techniques to Disentangle Confounding	230
7.3	Adjustment Methods Based on Stratification	234
7.4	Multiple Regression Techniques for Adjustment	248
7.5	Alternative Approaches for the Control of Confounding	282
7.6	Incomplete Adjustment: Residual Confounding	293
7.7	Over-Adjustment	295
7.8	Conclusion	296
	References	301
	Exercises	305

CHAPTER 8 **Quality Assurance and Control** **313**

8.1	Introduction	313
8.2	Quality Assurance	313
8.3	Quality Control	318
8.4	Indices of Validity and Reliability	328
8.5	Regression to the Mean	358
8.6	Final Considerations	359
	References	360
	Exercises	363

PART FIVE **Issues of Reporting and Application of Epidemiologic Results** **367**

CHAPTER 9 **Communicating Results of Epidemiologic Studies** **369**

9.1	Introduction	369
9.2	What to Report	369
9.3	How to Report	373
9.4	Conclusion	385
	References	386
	Exercises	388

CHAPTER 10 **Epidemiologic Issues in the Interface with Public Health Policy** **391**

10.1	Introduction	391
10.2	Causality: Application to Public Health and Health Policy	392
10.3	Decision Tree and Sensitivity Analysis	408
10.4	Meta-Analysis	413
10.5	Publication Bias	416
10.6	Summary	420
	References	421
	Exercises	426

APPENDIX A	Standard Errors, Confidence Intervals, and Hypothesis Testing for Selected Measures of Risk and Measures of Association	431
APPENDIX B	Test for Trend (Dose Response)	455
APPENDIX C	Test of Homogeneity of Stratified Estimates (Test for Interaction)	459
APPENDIX D	Quality Assurance and Quality Control Procedures Manual for Blood Pressure Measurement and Blood/Urine Collection in the ARIC Study	461
APPENDIX E	Calculation of the Intraclass Correlation Coefficient	469
APPENDIX F	Answers to Exercises	473
	Index	501

Preface

This book was conceived as an intermediate epidemiology textbook. Similarly to the first and second editions, the third edition explores and discusses key epidemiologic concepts and basic methods in more depth than that found in basic textbooks on epidemiology. For the third edition, new examples and exercises have been added to all chapters. In Chapters 7 and 10, respectively, we included discussions of novel epidemiologic strategies for handling confounding (i.e., instrumental variables and propensity scores) and of decision tree as a decision-making tool.

As an intermediate methods text, this book is expected to have a heterogeneous readership. Epidemiology students may wish to use it as a bridge between basic and more advanced epidemiologic methods. Other readers may desire to advance their knowledge beyond basic epidemiologic principles and methods but are not statistically minded and are thus reluctant to tackle the many excellent textbooks that strongly focus on epidemiology's quantitative aspects. The demonstration of several epidemiologic concepts and methods needs to rely on statistical formulations, and this text extensively supports these formulations with real-life examples, hopefully making their logic intuitively easier to follow. The practicing epidemiologist may find selected portions of this book useful for an understanding of concepts beyond the basics. Thus, the common denominators for the intended readers are familiarity with the basic strategies of analytic epidemiology and a desire to increase their level of understanding of several notions that are insufficiently covered (and naturally so) in many basic textbooks. The way in which this textbook is organized makes this readily apparent.

In Chapter 1, the basic observational epidemiologic research strategies are reviewed, including those based on studies of both groups and individuals. Although descriptive epidemiology is not the focus of this book, birth cohort analysis is discussed in some depth in this chapter because this approach is rarely covered in detail in basic textbooks. Another topic in the interface between descriptive and analytical epidemiology—namely, ecological studies—is also discussed, with a view toward extending its discussion beyond the possibility of inferential (ecological) bias. Next, the chapter reviews observational studies based on individuals as units of observation—that is, cohort and case-control studies. Different types of case-control design are reviewed. The strategy of *matching* as an approach by which to achieve comparability prior to data collection is also briefly discussed.

Chapters 2 and 3 cover issues of measurement of outcome frequency and measures of association. In Chapter 2, absolute measures of outcome frequency and their calculation methods are reviewed, including the person-time approach for the calculation of incidence density, and both the classic life table and the Kaplan-Meier method for the calculation of cumulative incidence. Chapter 3 deals with measures of association, including those based on relative (e.g., relative risk, odds ratio) and absolute (attributable risk) differences. The connections between measures of association obtained in

cohort and case-control studies are emphasized. In particular, a description is given of the different measures of association (i.e., odds ratio, relative risk, rate ratio) that can be obtained in case-control studies as a function of the control selection strategies that were introduced in Chapter 1.

Chapters 4 and 5 are devoted to threats to the validity of epidemiologic studies, namely bias and confounding. In Chapter 4, the most common types of bias are discussed, including selection bias and information bias. In the discussion of information bias, simple examples are given to improve the understanding of the phenomenon of misclassification resulting from less-than-perfect sensitivity and specificity of the approaches used for ascertaining exposure, outcome, and/or confounding variables. This chapter also provides a discussion of cross-sectional biases and biases associated with evaluation of screening procedures; for the latter, a simple approach to estimate lead time bias is given, which may be useful for those involved in evaluative studies of this sort. In Chapter 5, the concept of confounding is introduced, and approaches to evaluate confounding are reviewed. Special issues related to confounding are discussed, including the distinction between confounders and intermediate variables, residual confounding, and the role of statistical significance in the evaluation of confounding effects.

Interaction (effect modification) is discussed in Chapter 6. The chapter presents the concept of interaction, emphasizing its pragmatic application as well as the strategies used to evaluate the presence of additive and multiplicative interactions. Practical issues discussed in this chapter include whether to adjust when interaction is suspected and the importance of the additive model in public health.

The next three chapters are devoted to the approaches used to handle threats to the validity of epidemiologic results. In Chapter 7, strategies for the adjustment of confounding factors are presented, including the more parsimonious approaches (e.g., direct adjustment, Mantel-Haenszel) and the more complex (i.e., multiple regression, instrumental variables, Mendelian randomization, and propensity scores). Emphasis is placed on the selection of the method that is most appropriate for the study design used (e.g., Cox proportional hazards for the analysis of survival data or Poisson regression for the analysis of rates per person-time). Chapter 8 reviews the basic quality-control strategies for the prevention and control of measurement error and bias. Both qualitative and quantitative approaches used in quality control are discussed. The most-often used analytic strategies for estimating validity and reliability of data obtained in epidemiologic studies are reviewed (e.g., unweighted and weighted kappa, correlation coefficients) in this chapter. In Chapter 9, the key issue of communication of results of epidemiologic studies is discussed. Examples of common mistakes made when reporting epidemiologic data are given as a way to stress the importance of clarity in such reports.

Chapter 10 discusses—from the epidemiologist's viewpoint—issues relevant to the interface between epidemiology, health policy, and public health, such as Rothman's causality model, proximal and distal causes, and Hill's guidelines. This chapter also includes brief discussions of three topics pertinent to causal inference: sensitivity analysis, meta-analysis, and publication bias; and consideration of the decision tree as a tool to evaluate interventions. As in the previous editions, Appendices A, B, C, and E describe selected statistical procedures (e.g., standard errors and confidence levels, trend test, test of heterogeneity of effects, intraclass correlation) to help the reader to more thoroughly evaluate the measures of risk and association discussed in the

text and to expose him or her to procedures that, although relatively simple, are not available in many statistical packages used by epidemiology students and practitioners. Appendix D includes two sections on quality assurance and control procedures taken from the corresponding manual of the Atherosclerosis Risk in Communities (ARIC) Study as examples of real-life applications of some of the procedures discussed in Chapter 8. Finally, Appendix F provides the answers to the exercises.

We encourage readers to advise us of any errors or unclear passages, and to suggest improvements. Please email any such suggestions or comments to: info@jblearning.com.

All significant contributions will be acknowledged in the next edition.

Acknowledgments

This book is an outgrowth of an intermediate epidemiology course taught by the authors at the Johns Hopkins Bloomberg School of Public Health. Over the years, this course has benefited from significant intellectual input of many faculty members, including, among others, George W. Comstock, Helen Abbey, James Tonascia, Leon Gordis, and Mary Meyer. The authors especially acknowledge the late George W. Comstock, a mentor to both of us, who was involved with the course for several decades. His in-depth knowledge of epidemiologic methods and his wisdom over the years has been instrumental to our professional growth. Dr. Comstock also kindly provided many of the materials and examples used in Chapter 9 of this book.

We are indebted to many colleagues, including Leonelo Bautista, Daniel Brotman, Woody Chambless, Steve Cole, Joseph Coresh, Rosa Maria Corona, Ana Diez-Roux, Jingzhong Ding, Manning Feinleib, Leon Gordis, Eliseo Guallar, Jay Kaufman, Kristen Malecki, Alfonso Mele, Paolo Pasquini, Paul Peppard, Patrick Remington, Jonathan Samet, Eyal Shahar, Richey Sharrett, and Michael Silverberg. These colleagues reviewed partial sections of this or previous editions or provided guidance in solving conceptual or statistical riddles. We are especially grateful to our current and former students Gabrielle Detjen, Salwa Massad, Margarete (Grete) Wichmann and Hannah Yang for their careful review of the exercises. The authors are also grateful to Lauren Wisk for creating the ancillary instructor materials for this text. Finally, we would like to extend our appreciation to Patty Grubb, whose assistance has been instrumental in getting this edition ready, and to Jennifer Seltzer for her administrative help.

To have enjoyed the privilege of teaching intermediate epidemiology for so many years made us realize how much we have learned from our students, to whom we are deeply grateful. Finally, without the support and extraordinary patience of all members of our families, particularly our wives, Hilda and Marion, we could not have devoted so much time and effort to writing the three editions of this text.

About the Authors

Moyses Szklo, MD, DrPH, is a Professor of Epidemiology and Medicine (Cardiology) at the Johns Hopkins University. His current research focuses on risk factors for subclinical and clinical atherosclerosis. He is also Editor-in-Chief of the *American Journal of Epidemiology*.

E. Javier Nieto, MD, PhD, is Helfaer Professor of Public Health, and Professor and Chair at the Department of Population Health Sciences at the University of Wisconsin School of Medicine and Public Health. His current research focuses on epidemiology of cardiovascular and sleep disorders as well as on population-based survey methods.

