

## CAUSAL INFERENCE

---

### **for Statistics, Social, and Biomedical Sciences An Introduction**

Most questions in social and biomedical sciences are causal in nature: what would happen to individuals, or to groups, if part of their environment were changed? In this groundbreaking text, two world-renowned experts present statistical methods for studying such questions.

This book starts with the notion of potential outcomes, each corresponding to the outcome that would be realized if a subject were exposed to a particular treatment or regime. In this approach, causal effects are comparisons of such potential outcomes. The fundamental problem of causal inference is that we can observe only one of the potential outcomes for a particular subject. The authors discuss how randomized experiments allow us to assess causal effects and then turn to observational studies. They lay out the assumptions needed for causal inference and describe the leading analysis methods, including matching, propensity-score methods, and instrumental variables. Many detailed applications are included, with special focus on practical aspects for the empirical researcher.

Guido W. Imbens is Applied Econometrics Professor and Professor of Economics at the Graduate School of Business at Stanford University. He has previously held tenured positions at the University of California at Los Angeles, the University of California at Berkeley, and Harvard University. He is a Fellow of the Econometric Society and the American Academy of Arts and Sciences. He holds an honorary doctorate from the University of St. Gallen, Switzerland. Imbens has done extensive research in econometrics and statistics, specializing in causal inference. He has published widely in leading economics and statistics journals, including the *American Economic Review*, *Econometrica*, the *Review of Economic Studies*, the *Journal of the American Statistical Association*, the *Annals of Statistics*, *Biometrika*, and the *Journal of the Royal Statistical Society, Series A*.

Donald B. Rubin is John L. Loeb Professor of Statistics at Harvard University, where he has been professor since 1983 and Department chair for 13 of those years. He has been elected to be a Fellow/Member/Honorary Member of the Woodrow Wilson Society, the Guggenheim Memorial Foundation, the Alexander von Humboldt Foundation, the American Statistical Association, the Institute of Mathematical Statistics, the International Statistical Institute, the American Association for the Advancement of Science, the American Academy of Arts and Sciences, the European Association of Methodology, the British Academy, and the U.S. National Academy of Sciences. As of 2014, he has authored/coauthored nearly 400 publications (including ten books), has five joint patents, and for many years has been one of the most highly cited authors in the world, with currently over 150,000 citations (Google Scholar). He is also the recipient of honorary doctorates from Otto Friedrich University, Bamberg, Germany; the University of Ljubljana, Slovenia; and Universidad Santa Tomás, Bogotá, Colombia, and has been named an honorary professor at four universities.

**Advance Praise for *Causal Inference for Statistics, Social, and Biomedical Sciences***

“This thorough and comprehensive book uses the ‘potential outcomes’ approach to connect the breadth of theory of causal inference to the real-world analyses that are the foundation of evidence-based decision making in medicine, public policy, and many other fields. Imbens and Rubin provide unprecedented guidance for designing research on causal relationships, and for interpreting the results of that research appropriately.”

– *Dr. Mark McClellan, Director of the Health Care Innovation and Value Initiative, the Brookings Institution*

“Clarity of thinking about causality is of central importance in financial decision making. Imbens and Rubin provide a rigorous foundation allowing practitioners to learn from the pioneers in the field.”

– *Dr. Stephen Blyth, Managing Director, Head of Public Markets, Harvard Management Company*

“A masterful account of the potential outcomes approach to causal inference from observational studies that Rubin has been developing since he pioneered it 40 years ago.”

– *Adrian Raftery, Blumstein-Jordan Professor of Statistics and Sociology, University of Washington*

“Correctly drawing causal inferences is critical in many important applications. Congratulations to Professors Imbens and Rubin, who have drawn on their decades of research in this area, along with the work of several others, to produce this impressive book covering concepts, theory, methods, and applications. I especially appreciate their clear exposition of conceptual issues, which are important to understand in the context of either a designed experiment or an observational study, and their use of real applications to motivate the methods described.”

– *Nathaniel Schenker, Former President of the American Statistical Association*

# CAUSAL INFERENCE

---

**for Statistics, Social, and  
Biomedical Sciences  
An Introduction**

**Guido W. Imbens**

*Stanford University*

**Donald B. Rubin**

*Harvard University*

**CAMBRIDGE**  
UNIVERSITY PRESS

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521885881](http://www.cambridge.org/9780521885881)

© Guido W. Imbens and Donald B. Rubin 2015

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2015

Reprinted 2016

Printed in the United States of America by Sheridan Books, Inc.

*A catalog record for this publication is available from the British Library.*

*Library of Congress Cataloging in Publication Data*

Imbens, Guido.

Causal inference for statistics, social, and biomedical sciences :  
an introduction / Guido W. Imbens & Donald B. Rubin.

pages cm

1. Social sciences—Research. 2. Causation. 3. Inference. I. Rubin, Donald B. II. Title.

H62.I537 2014

519.5'4—dc23 2014020988

ISBN 978-0-521-88588-1 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

*To my parents, Annie Imbens-Fransen and Gerard Imbens, for all their support and encouragement over the years; to my children, Carleton, Annalise and Sylvia, who have provided so much joy in recent years; and to my wife and best friend, Susan Athey.*

*Guido W. Imbens*

*To my family, colleagues, and students, and Elizabeth Zell, who has provided ten years of crucial support.*

*Donald B. Rubin*

## Contents

<i>Preface</i>	<i>page xvii</i>
PART I INTRODUCTION	
<b>1 Causality: The Basic Framework</b>	<b>3</b>
1.1 Introduction	3
1.2 Potential Outcomes	3
1.3 Definition of Causal Effects	5
1.4 Causal Effects in Common Usage	7
1.5 Learning about Causal Effects: Multiple Units	8
1.6 The Stable Unit Treatment Value Assumption	9
1.7 The Assignment Mechanism: An Introduction	13
1.8 Attributes, Pre-Treatment Variables, or Covariates	15
1.9 Potential Outcomes and Lord's Paradox	16
1.10 Causal Estimands	18
1.11 Structure of the Book	20
1.12 Samples, Populations, and Super-Populations	20
1.13 Conclusion	21
Notes	21
<b>2 A Brief History of the Potential Outcomes Approach to Causal Inference</b>	<b>23</b>
2.1 Introduction	23
2.2 Potential Outcomes and the Assignment Mechanism before Neyman	24
2.3 Neyman's (1923) Potential Outcome Notation in Randomized Experiments	25
2.4 Earlier Hints for Physical Randomizing	26
2.5 Fisher's (1925) Proposal to Randomize Treatments to Units	26
2.6 The Observed Outcome Notation in Observational Studies for Causal Effects	27
2.7 Early Uses of Potential Outcomes in Observational Studies in Social Sciences	28
	vii

2.8	Potential Outcomes and the Assignment Mechanism in Observational Studies: Rubin (1974) Notes	29 30
<b>3</b>	<b>A Classification of Assignment Mechanisms</b>	<b>31</b>
3.1	Introduction	31
3.2	Notation	33
3.3	Assignment Probabilities	34
3.4	Restrictions on the Assignment Mechanism	37
3.5	Assignment Mechanisms and Super-Populations	39
3.6	Randomized Experiments	40
3.7	Observational Studies: Regular Assignment Mechanisms	41
3.8	Observational Studies: Irregular Assignment Mechanisms	42
3.9	Conclusion Notes	43 43
PART II CLASSICAL RANDOMIZED EXPERIMENTS		
<b>4</b>	<b>A Taxonomy of Classical Randomized Experiments</b>	<b>47</b>
4.1	Introduction	47
4.2	Notation	48
4.3	Bernoulli Trials	48
4.4	Completely Randomized Experiments	50
4.5	Stratified Randomized Experiments	51
4.6	Paired Randomized Experiments	52
4.7	Discussion	53
4.8	Conclusion Notes	55 56
<b>5</b>	<b>Fisher's Exact P-Values for Completely Randomized Experiments</b>	<b>57</b>
5.1	Introduction	57
5.2	The Paul et al. Honey Experiment Data	59
5.3	A Simple Example with Six Units	59
5.4	The Choice of Null Hypothesis	63
5.5	The Choice of Statistic	64
5.6	A Small Simulation Study	72
5.7	Interval Estimates Based on Fisher P-Value Calculations	74
5.8	Computation of P-Values	75
5.9	Fisher Exact P-Values with Covariates	78
5.10	Fisher Exact P-Values for the Honey Data	80
5.11	Conclusion Notes	81 81
<b>6</b>	<b>Neyman's Repeated Sampling Approach to Completely Randomized Experiments</b>	<b>83</b>
6.1	Introduction	83
6.2	The Duflo-Hanna-Ryan Teacher-Incentive Experiment Data	84
6.3	Unbiased Estimation of the Average Treatment Effect	85

<b>Contents</b>	ix
6.4 The Sampling Variance of the Neyman Estimator	87
6.5 Estimating the Sampling Variance	92
6.6 Confidence Intervals and Testing	95
6.7 Inference for Population Average Treatment Effects	98
6.8 Neyman's Approach with Covariates	101
6.9 Results for the Duflo-Hanna-Ryan Teacher-Incentive Data	102
6.10 Conclusion	104
Notes	104
Appendix A Sampling Variance Calculations	105
Appendix B Random Sampling from a Super-Population	109
<b>7 Regression Methods for Completely Randomized Experiments</b>	113
7.1 Introduction	113
7.2 The LRC-CPPT Cholesterol Data	115
7.3 The Super-Population Average Treatment Effects	116
7.4 Linear Regression with No Covariates	118
7.5 Linear Regression with Additional Covariates	122
7.6 Linear Regression with Covariates and Interactions	125
7.7 Transformations of the Outcome Variable	127
7.8 The Limits on Increases in Precision Due to Covariates	128
7.9 Testing for the Presence of Treatment Effects	129
7.10 Estimates for LRC-CPPT Cholesterol Data	131
7.11 Conclusion	133
Notes	134
Appendix	135
<b>8 Model-Based Inference for Completely Randomized Experiments</b>	141
8.1 Introduction	141
8.2 The Lalonde NSW Experimental Job-Training Data	144
8.3 A Simple Example: Naive and More Sophisticated Approaches to Imputation	146
8.4 Bayesian Model-Based Imputation in the Absence of Covariates	150
8.5 Simulation Methods in the Model-Based Approach	163
8.6 Dependence between Potential Outcomes	165
8.7 Model-Based Imputation with Covariates	169
8.8 Super-Population Average Treatment Effects	171
8.9 A Frequentist Perspective	172
8.10 Model-Based Estimates of the Effect of the NSW Program	174
8.11 Conclusion	177
Notes	177
Appendix A Posterior Distributions for Normal Models	178
Appendix B Analytic Derivations with Known Covariance Matrix	181
<b>9 Stratified Randomized Experiments</b>	187
9.1 Introduction	187
9.2 The Tennessee Project Star Data	188



9.3	The Structure of Stratified Randomized Experiments	189
9.4	Fisher's Exact P-Values in Stratified Randomized Experiments	192
9.5	The Analysis of Stratified Randomized Experiments from Neyman's Repeated Sampling Perspective	201
9.6	Regression Analysis of Stratified Randomized Experiments	205
9.7	Model-Based Analysis of Stratified Randomized Experiments	207
9.8	Design Issues: Stratified versus Completely Randomized Experiments	211
9.9	Conclusion	212
	Notes	212
	Appendix A: Student-Level Analyses	213
	Appendix B: Proofs of Theorems 9.1 and 9.2	214
<b>10</b>	<b>Pairwise Randomized Experiments</b>	219
10.1	Introduction	219
10.2	The Children's Television Workshop Experiment Data	220
10.3	Pairwise Randomized Experiments	220
10.4	Fisher's Exact P-Values in Pairwise Randomized Experiments	222
10.5	The Analysis of Pairwise Randomized Experiments from Neyman's Repeated Sampling Perspective	224
10.6	Regression-Based Analysis of Pairwise Randomized Experiments	229
10.7	Model-Based Analysis of Pairwise Randomized Experiments	231
10.8	Conclusion	233
	Notes	234
	Appendix: Proofs	234
<b>11</b>	<b>Case Study: An Experimental Evaluation of a Labor Market Program</b>	240
11.1	Introduction	240
11.2	The San Diego SWIM Program Data	240
11.3	Fisher's Exact P-Values	242
11.4	Neyman's Repeated Sampling-Based Point Estimates and Large-Sample Confidence Intervals	245
11.5	Regression-Based Estimates	247
11.6	Model-Based Point Estimates	250
11.7	Conclusion	253
	Notes	253
PART III REGULAR ASSIGNMENT MECHANISMS: DESIGN		
<b>12</b>	<b>Unconfounded Treatment Assignment</b>	257
12.1	Introduction	257
12.2	Regular Assignment Mechanisms	258
12.3	Balancing Scores and the Propensity Score	266
12.4	Estimation and Inference	268
12.5	Design Phase	276

<b>Contents</b>	xi
12.6 Assessing Unconfoundedness	278
12.7 Conclusion	279
Notes	279
<b>13 Estimating the Propensity Score</b>	<b>281</b>
13.1 Introduction	281
13.2 The Reinisch et al. Barbiturate Exposure Data	284
13.3 Selecting the Covariates and Interactions	285
13.4 Choosing the Specification of the Propensity Score for the Barbiturate Data	288
13.5 Constructing Propensity-Score Strata	290
13.6 Choosing Strata for the Barbiturate Data	294
13.7 Assessing Balance Conditional on the Estimated Propensity Score	296
13.8 Assessing Covariate Balance for the Barbiturate Data	300
13.9 Conclusion	306
Notes	306
Appendix: Logistic Regression	307
<b>14 Assessing Overlap in Covariate Distributions</b>	<b>309</b>
14.1 Introduction	309
14.2 Assessing Balance in Univariate Distributions	310
14.3 Direct Assessment of Balance in Multivariate Distributions	313
14.4 Assessing Balance in Multivariate Distributions Using the Propensity Score	314
14.5 Assessing the Ability to Adjust for Differences in Covariates by Treatment Status	317
14.6 Assessing Balance: Four Illustrations	318
14.7 Sensitivity of Regression Estimates to Lack of Overlap	332
14.8 Conclusion	336
Notes	336
<b>15 Matching to Improve Balance in Covariate Distributions</b>	<b>337</b>
15.1 Introduction	337
15.2 The Reinisch et al. Barbiturate Exposure Data	339
15.3 Selecting a Subsample of Controls through Matching to Improve Balance	339
15.4 An Illustration of Propensity Score Matching with Six Observations	344
15.5 Theoretical Properties of Matching Procedures	345
15.6 Creating Matched Samples for the Barbiturate Data	349
15.7 Conclusion	358
Notes	358
<b>16 Trimming to Improve Balance in Covariate Distributions</b>	<b>359</b>
16.1 Introduction	359
16.2 The Right Heart Catheterization Data	360
16.3 An Example with a Single Binary Covariate	362

16.4	Selecting a Subsample Based on the Propensity Score	366
16.5	The Optimal Subsample for the Right Heart Catheterization Data	368
16.6	Conclusion	373
	Notes	374
PART IV REGULAR ASSIGNMENT MECHANISMS: ANALYSIS		
<b>17</b>	<b>Subclassification on the Propensity Score</b>	377
17.1	Introduction	377
17.2	The Imbens-Rubin-Sacerdote Lottery Data	378
17.3	Subclassification on the Propensity Score and Bias Reduction	380
17.4	Subclassification and the Lottery Data	385
17.5	Estimation Based on Subclassification with Additional Bias Reduction	386
17.6	Neymanian Inference	388
17.7	Average Treatment Effects for the Lottery Data	390
17.8	Weighting Estimators and Subclassification	392
17.9	Conclusion	399
	Notes	399
<b>18</b>	<b>Matching Estimators</b>	401
18.1	Introduction	401
18.2	The Card-Krueger New Jersey and Pennsylvania Minimum Wage Data	404
18.3	Exact Matching without Replacement	405
18.4	Inexact Matching without Replacement	407
18.5	Distance Measures	410
18.6	Matching and the Card-Krueger Data	412
18.7	The Bias of Matching Estimators	415
18.8	Bias-Corrected Matching Estimators	416
18.9	Matching with Replacement	424
18.10	The Number of Matches	425
18.11	Matching Estimators for the Average Treatment Effect for the Controls and for the Full Sample	427
18.12	Matching Estimates of the Effect of the Minimum Wage Increase	428
18.13	Conclusion	430
	Notes	431
<b>19</b>	<b>A General Method for Estimating Sampling Variances for Standard Estimators for Average Causal Effects</b>	433
19.1	Introduction	433
19.2	The Imbens-Rubin-Sacerdote Lottery Data	435
19.3	Estimands	436

<b>Contents</b>	xiii
19.4 The Common Structure of Standard Estimators for Average Treatment Effects	441
19.5 A General Formula for the Conditional Sampling Variance	445
19.6 A Simple Estimator for the Unit-Level Conditional Sampling Variance	446
19.7 An Estimator for the Sampling Variance of $\hat{\tau}$ Conditional on Covariates	452
19.8 An Estimator for the Sampling Variance for the Estimator for the Average Effect for the Treated	452
19.9 An Estimator for the Sampling Variance for the Population Average Treatment Effect	454
19.10 Alternative Estimators for the Sampling Variance	456
19.11 Conclusion	460
Notes	460
<b>20 Inference for General Causal Estimands</b>	461
20.1 Introduction	461
20.2 The Lalonde NSW Observational Job-Training Data	462
20.3 Causal Estimands	465
20.4 A Model for the Conditional Potential Outcome Distributions	468
20.5 Implementation	472
20.6 Results for the Lalonde Data	473
20.7 Conclusion	474
Notes	474
PART V REGULAR ASSIGNMENT MECHANISMS: SUPPLEMENTARY ANALYSES	
<b>21 Assessing Unconfoundedness</b>	479
21.1 Introduction	479
21.2 Setup	482
21.3 Estimating Effects on Pseudo-Outcomes	482
21.4 Estimating Effects of Pseudo-Treatments	485
21.5 Robustness to the Set of Pre-Treatment Variables	487
21.6 The Imbens-Rubin-Sacerdote Lottery Data	490
21.7 Conclusion	495
Notes	495
<b>22 Sensitivity Analysis and Bounds</b>	496
22.1 Introduction	496
22.2 The Imbens-Rubin-Sacerdote Lottery Data	497
22.3 Bounds	497
22.4 Binary Outcomes: The Rosenbaum-Rubin Sensitivity Analysis	500

22.5	Binary Outcomes: The Rosenbaum Sensitivity Analysis for P-Values	506
22.6	Conclusion	509
	Notes	509
PART VI REGULAR ASSIGNMENT MECHANISMS WITH NONCOMPLIANCE: ANALYSIS		
<b>23</b>	<b>Instrumental Variables Analysis of Randomized Experiments with One-Sided Noncompliance</b>	513
23.1	Introduction	513
23.2	The Sommer-Zeger Vitamin A Supplement Data	516
23.3	Setup	517
23.4	Intention-to-Treat Effects	519
23.5	Compliance Status	522
23.6	Instrumental Variables	526
23.7	Moment-Based Instrumental Variables Estimators	530
23.8	Linear Models and Instrumental Variables	531
23.9	Naive Analyses: “As-Treated,” “Per Protocol,” and Unconfoundedness	535
23.10	Conclusion	539
	Notes	539
	Appendix	541
<b>24</b>	<b>Instrumental Variables Analysis of Randomized Experiments with Two-Sided Noncompliance</b>	542
24.1	Introduction	542
24.2	The Angrist Draft Lottery Data	543
24.3	Compliance Status	544
24.4	Intention-to-Treat Effects	546
24.5	Instrumental Variables	548
24.6	Traditional Econometric Methods for Instrumental Variables	556
24.7	Conclusion	559
	Notes	559
<b>25</b>	<b>Model-Based Analysis in Instrumental Variable Settings: Randomized Experiments with Two-Sided Noncompliance</b>	560
25.1	Introduction	560
25.2	The McDonald-Hiu-Tierney Influenza Vaccination Data	561
25.3	Covariates	567
25.4	Model-Based Instrumental Variables Analyses for Randomized Experiments with Two-Sided Noncompliance	568
25.5	Simulation Methods for Obtaining Draws from the Posterior Distribution of the Estimand Given the Data	574
25.6	Models for the Influenza Vaccination Data	578

<b>Contents</b>	xv
25.7 Results for the Influenza Vaccination Data	581
25.8 Conclusion	584
Notes	584
PART VII CONCLUSION	
<b>26 Conclusions and Extensions</b>	589
Notes	590
<i>References</i>	591
<i>Author Index</i>	605
<i>Subject Index</i>	609

## Preface

In many applications of statistics, a large proportion of the questions of interest are fundamentally questions of causality rather than simply questions of description or association. For example, a medical researcher may wish to find out whether a new drug is effective against a disease. An economist may be interested in uncovering the effects of a job-training program on an individual's employment prospects, or the effects of a new tax or regulation on economic activity. A sociologist may be concerned about the effects of divorce on children's subsequent education. In this text we discuss statistical methods for studying such questions.

The book arose out of a conversation we had in 1992 while we were both on the faculty at Harvard University. We found that although we were both interested in questions of causality, we had difficulty communicating our ideas because, coming from different disciplines, we were used to different terminology and conventions. However, the excitement about the ideas in these different areas motivated us to capitalize on these difficulties, which led to a long collaboration, including research projects, graduate and undergraduate teaching, and thesis advising. The book is a reflection of this collaboration.

The book is based directly on many semester and quarter-length courses we, initially jointly, and later separately, taught for a number of years, starting in 1995 at Harvard University, followed by the University of California at Los Angeles, the University of California at Berkeley, and Stanford University, to audiences of graduate and undergraduate students from statistics, economics, business, and other disciplines using applied statistics. In addition, we have taught shorter versions of such courses in Barcelona, Beijing, Berlin, Bern, Geneva, Maastricht, Mexico City, Miami, Montevideo, Santiago, Stockholm, Uppsala, Wuppertal, Zurich, and at the World Bank as well as other associations and agencies.

There are a number of key features of the approach taken in this book. First of all, the perspective we take is that all causal questions are tied to specific interventions or treatments. Second, causal questions are viewed as comparisons of potential outcomes, with each potential outcome corresponding to a level of the treatment. Each of these potential outcomes could have been observed had the treatment taken on the corresponding level. After the treatment has taken on a specific level, only the potential

outcome corresponding to that level is realized and can actually be observed. Causal effects involve the comparison of the outcome actually observed with other potential outcomes that could have been observed had the treatment taken on a different level, but that are not, in fact, observed. Causal inference is therefore fundamentally a missing data problem and, as in all missing data problems, a key role is played by the mechanism that determines which data values are observed and which are missing. In causal inference, this mechanism is referred to as the assignment mechanism, the mechanism that determines levels of the treatment taken by the units studied.

The book is organized in seven parts. In the first part we set out the basic philosophy underlying our approach to causal inference and describe the potential outcomes framework. The next three parts of the book are distinguished by the assumptions maintained about the assignment mechanism. In Part II we assume that the assignment mechanism corresponds to a classical randomized experiment.

In Part III we assume that the assignment mechanism is “regular” in a well-defined sense, which generalizes randomized experiments. In this part of the book we discuss what we call the “design” phase of an observational study, which we view as extremely important for credible conclusions. In the next part, Part IV, we discuss data analysis for studies with regular assignment mechanisms. Here we consider matching and subclassification procedures, as well as model-based and weighting methods.

In Part V we relax this regularity assumption and discuss more general assignment mechanisms. First we assess the key assumption required for regularity, unconfoundedness. We also explore in this part of the text sensitivity analyses where we relax some of the key features of a regular assignment mechanism.

Next, in Part VI of the text, we consider settings where the assignment mechanism is regular, but compliance with the assignment is imperfect. As a result, the probability of receipt of treatment may depend on both observed and unobserved characteristics and outcomes of the units. To address these complications, we turn to instrumental variables methods. Part VII of the book concludes.

As with all books, ours has limitations. Foremost is our focus on binary treatments. Although many of the results can easily be extended to multi-valued treatments, we focus on the binary treatment case because many critical conceptual issues arise already in that setting. Second, throughout most of the book we make the “stability” assumption that treatments applied to one unit do not affect outcomes for other units and that there are no unrepresented versions of the treatments. There is a growing literature on interactions through networks and peer effects that builds on the notions of causality discussed in this book. Finally, although we designed the book to be theoretically tight and principled, we focus on practical rather than mathematical results, including detailed applications with real data sets, consistent with our target audience of researchers in applied fields.

## ACKNOWLEDGMENTS

We are deeply indebted to many collaborators, colleagues, and students from whom we have learned much about the topics discussed in this book over the years. These include Alberto Abadie, Nikola Andric, Joshua Angrist, Susan Athey, Thomas Barrios, David Card, Matias Cattaneo, Gary Chamberlain, Raj Chetty, Rajeev Dehejia, Mike Dickstein,



**Preface**

xix

Peng Ding, David Drukker, Dean Eckles, Valeria Espinosa, Avi Feller, Sergio Firpo, Andrew Gelman, Paul Gift, Paul Goldsmith-Pinkham, Bryan Graham, Roe Gutman, Gary Harvey, Jane Herr, Kei Hirano, Nate Holt, Joe Hotz, Wilbert van der Klaauw, Jacob Klerman, Alan Krueger, Tony Lancaster, Joseph Lee, Victoria Liublinska, Chuck Manski, Eric Maskin, Fabrizia Mealli, Kari Lock Morgan, Eduardo Morales, Julie Mortimer, Eleanor Murray, Whitney Newey, Cassandra Wolos Pattanayak, Jack Porter, Geert Ridder, Paul Rosenbaum, Bruce Sacerdote, Hal Stern, Elizabeth Stuart, Neal Thomas, Sadek Wahba, David Watson, Jeff Wooldridge, Alan Zaslavsky, and Elizabeth Zell.

We are grateful for stimulating discussions over the years with David Freedman, Sander Greenland, James Heckman, Nick Jewell, Mark van der Laan, Judea Pearl, Jamie Robins, Jasjeet Sekhon, Jeffrey Smith, and Edward Vytlačil.

We are also grateful to Joshua Angrist, David Card, Raj Chetty, Rajeev Dehejia, Esther Duflo, Brad Efron, Alan Krueger, Robert Lalonde, June Reinisch, Bruce Sacerdote, Sadek Wahba, and Scott Zeger for making their data available.

Finally, we would like to extend our gratitude to Dean Eckles and Karthik Rajkumar for comments on the first edition, and to Michael Pollman for a remarkably thorough check for errors and typos.