Medical Uses of Statistics

THIRD EDITION

EDITED BY

JOHN C. BAILAR III

DAVID C. HOAGLIN

WILEY

The NEW ENGLAND JOURNAL of MEDICINE

Medical Uses of Statistics

3rd Edition

Medical Uses of Statistics

3rd Edition

The NEW ENGLAND JOURNAL of MEDICINE

EDITED BY

JOHN C. BAILAR III
DAVID C. HOAGLIN

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada

Library of Congress Cataloging in Publication Data:
Medical uses of statistics / edited by John C. Bailar III, David C. Hoaglin. — 3rd ed.
p. ; cm.
Includes articles originally published in the New England journal of medicine.
Includes bibliographical references and index.
ISBN 978-0-470-43952-4 (cloth) — ISBN 978-0-470-43953-1 (pbk.)
1. Medical statistics. 2. Clinical medicine — Research — Statistical methods. I. Bailar III, John C. (John Christian), 1932- II. Hoaglin, David C. (David Caster), 1944- III. New England Journal of Medicine.
[DNLM: 1. Statistics as Topic — Collected Works. 2. Research — methods — Collected Works. WA 950 M489 2009]
RA409.M43 2009
610.72—dc22

2009017256

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

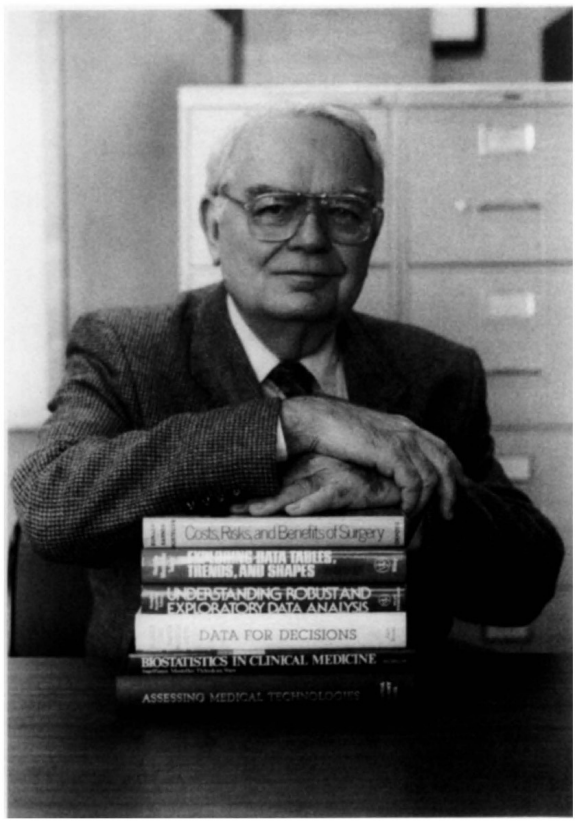
To

Frederick Mosteller (1916–2006)

superb teacher

supportive friend

and wise collaborator



Contents

Contributors

Shilpi Agarwal, M.B.B.S.
Department of Epidemiology; Harvard School of Public Health

Paul S. Albert, Ph.D.
Biometric Research Branch, National Cancer Institute

John C. Bailar III, M.D., Ph.D.
Professor Emeritus, University of Chicago; Scholar in Residence, National Academies

A. John Bailer, Ph.D.
Department of Mathematics & Statistics, Miami University

Graham A. Colditz, M.D., Dr.P.H.
Department of Surgery, Washington University School of Medicine

Fernando Delgado, M.S.
Colombia, South America

Christl Donnelly, D.Sc.
Department of Biostatistics, School of Public Health, Harvard University

Jeffrey M. Drazen, M.D.
Editor-in-Chief, New England Journal of Medicine

John D. Emerson, Ph.D.
Department of Mathematics, Middlebury College

Mark S. Goldberg, Ph.D.
Department of Medicine, McGill University

David C. Hoaglin, Ph.D.
Abt Bio-Pharma Solutions, Inc.

Hossein Hosseini, Ph.D.
Digital Equipment Corporation, Irvine, California

David J. Hunter, M.B.B.S.
Department of Medicine, Brigham & Women's Hospital; Harvard School of Public Health

Joseph A. Ingelfinger, M.D.
Bowdoin Street Health Center, Harvard Medical School

Thorsten Kurz, Ph.D.
Core Facility Genomics, University Hospital Freiburg, Germany

Stephen W. Lagakos, Ph.D.
Department of Biostatistics, School of Public Health, Harvard University

Philip W. Lavori, Ph.D.
Department of Psychiatry and Human Behavior, Brown University

Thomas A. Louis, Ph.D.
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Nancy E. Mayo, Ph.D.
Division of Clinical Epidemiology, Department of Medicine, McGill University

Stephen Morrissey, Ph.D.
New England Journal of Medicine

Lincoln E. Moses, Ph.D. (1921–2006)
Department of Statistics, Stanford University

Frederick Mosteller, Ph.D. (1916–2006)
Department of Statistics, Harvard University

Dan L. Nicolae, Ph.D.
Department of Medicine and Department of Statistics, University of Chicago

Carole Ober, Ph.D.
Department of Human Genetics, University of Chicago

Margaret Perkins, M.A.
New England Journal of Medicine

Marcia Polansky, D.Sc.,
Department of Biometrics and Computing, Drexel University

Amita Rastogi, M.D., M.H.A.
Ingenix, Inc.

Paul J. Rathouz, Ph.D.
Department of Health Studies, University of Chicago

Michael A. Stoto, Ph.D.
School of Nursing & Health Studies, Georgetown University

Rui Wang, Ph.D.
Biostatistics Center, Massachusetts General Hospital; Department of Biostatistics, Harvard School of Public Health

James H. Ware, Ph.D.
Department of Biostatistics, School of Public Health, Harvard University

Preface

The practice of medicine combines science and art. The science part of medicine derives largely from inferences drawn from experiments, often performed with the invaluable assistance of patients who put themselves at risk to become research participants. These brave and altruistic people have all or part of their medical care driven by the requirements of research participation rather than by their specific clinical needs. Investigators measure various outcomes and assemble the results of their observations in research reports, which medical journals review and publish to help guide the community's thinking about how best to approach the biology, prevention, diagnosis, and treatment of the condition under study.

It comes as no surprise that the clinical and laboratory observations involve many sources of variation, including measurement errors, intrinsic patient biological variability, and differences among patients in adherence to treatment protocols. These multiple sources of variation lead to uncertainty in assessments of outcome and in the clinical inferences drawn from them. Medical researchers apply statistical methods to these inherently noisy data and derive reasonably precise conclusions from them, taking into account not only the uncertainty but also other limitations of the data. Their experience with this process and its results also guides them in designing new studies. The conclusions drawn from these inferences drive clinical practice.

This third edition of *Medical Uses of Statistics* provides a broad first course in understanding the key ideas of quantitative methods that guide this process. Because we are interested in helping people understand the approaches used to study and solve problems rather than in providing a detailed manual for the investigator, concepts are explained with minimal use of mathematics. The approach maintains the emphasis in the first two editions, but this edition has been updated to include new methods and new disciplines. In the 17 years since publication of the second edition, new methods such as those used in genome-wide association studies or in multiple imputation for missing data have come into common use in medical journals. Because medicine is taught by example, the authors include multiple examples drawn from published articles, particularly from the *New England Journal of Medicine*, to illustrate each of the approaches and keep the presentation on firm practical ground. For the novice the book outlines the major statistical approaches used in medical analysis; for the expert the examples can provide hints about optimal study design and improvements in reporting results.

Regardless of your prior experience and expertise, it is highly likely, $p<0.001$, that this book will be a useful companion in the search for better information to guide clinical thinking. You can bet on it—keep reading, and you will see.

Jeffrey M. Drazen, M.D.
Editor-in-Chief, *New England Journal of Medicine*

Preface to the Second Edition (1992)

The first edition of this book, published over five years ago, found favor with a gratifyingly large number of readers and was widely praised as a unique contribution to its field. The Preface to the first edition, reprinted in almost its entirety, describes the book's origins and purposes. This second edition builds on the strengths of the first, extending its scope to new topics, while revising and updating treatment of many of the old ones and replacing a few of the original chapters with entirely new material.

The result is a slightly longer book, but I believe it is even better and more useful than its predecessor. The general philosophy and organization remain the same, but the range of subjects is broader and the overall treatment more comprehensive. Every effort has been made to achieve a readable and interesting text that explains the important ideas behind current medical uses of statistics without burdening the reader with the technical details of mathematical manipulations.

I found this new edition more interesting and accessible than the first. I trust readers will enjoy it as much as I did.

Arnold S. Relman, M.D.
Editor-in-Chief Emeritus, *New England Journal of Medicine*

*Preface to the First Edition (1986)**

No one who reads the current medical literature, and certainly no one who performs clinical studies these days, can be unaware of the growing importance of statistics. Sound clinical research, as well as the ability to understand published results of research, increasingly depends on a clear comprehension of the fundamental concepts of statistical design and analysis.

This book is the fruit of an idea that originated in 1977, in conversations with John Bailar and Frederick Mosteller of the Department of Biostatistics of the Harvard School of Public Health. Convinced that the readers of the *New England Journal of Medicine* needed a clearer idea of how statistical techniques were being applied in current clinical studies, my editorial colleagues and I (including most prominently our former Deputy Editor, Dr. Drummond Rennie) suggested to Bailar and Mosteller that they organize a study of the research papers published in recent volumes of the *Journal* (and some other important medical journals), to determine what statistical methods were actually being used. We also asked them to tell us whether the methods were appropriately applied and how their use might be improved, and we asked them to do so in simple language that would be understood even by readers who had no education in biostatistics.

With the aid of a generous grant from the Rockefeller Foundation, Bailar and Mosteller, assisted by a host of colleagues at Harvard and elsewhere, set out to do just that. Their work was greatly helped by encouragement from Dr. Kenneth Warren, Director of the Division of Health Sciences, and Dr. Kerr White, Special Projects Officer at the Rockefeller Foundation.

The result, in my view, has been spectacular. First of all, they carried out a survey of statistical practice in the *New England Journal* and a few other journals, demonstrating the frequency with which different types of statistical methods were applied and identifying the need for improvement in the selection and use of these methods. In addition, the group produced a series of articles on a wide range of statistical subjects, drawn from the insights gained during their survey of actual practice.

All together, more than 30 papers have come from this project so far. Some have appeared in the *Journal* as part of our "Statistics in Practice" series. A dozen or so have been published in other journals or as book chapters. Still others have been reserved for first publication in this book.

*Text appears as published in the second edition.

There are many books on biostatistics, but there are two unique and important characteristics of this one that I believe set it apart. First of all, as already noted, it is based on current usage, and it is concerned with improving that usage. Unlike most standard textbooks, this book takes an empirical, practical approach. It does not simply use examples from the literature to illustrate didactic points; it carefully surveys what clinical investigators are actually doing with statistical methods, as revealed mostly in the pages of the *Journal*. It tells readers what they need to know to understand those methods, and it points out ways in which medical writers can make their reporting of methods and results more informative and their analyses of data more useful.

Secondly, the orientation of this book is toward an understanding of ideas—when and why to use certain statistical techniques. There are many textbooks that explain statistical calculations but few or none that attempt, as this one does, to get behind the calculations and tell what they are all about. This book does not concern itself with the mechanics of statistical computation. There are no instructions on how to perform calculations, and there are few mathematical formulas. The emphasis here is on explaining the purpose of the statistical methods, so that the general reader will have a better understanding of the strategy to be employed and the alternatives that need to be considered. Most chapters, however, cite other "how-to" textbooks of statistics, to which readers may refer for detailed explanations of the mathematical calculations.

The authors have striven to write in a straightforward style, as unencumbered by biostatistical jargon as possible. Their object has been to make this book understandable to almost anyone who has a nodding acquaintance with biomedical research and an elementary grasp of numerical concepts. How well they have succeeded only the reader can judge, but, as an amateur myself, I have found their writing lucid and readable. I should think that most medical students and physicians—even those with no formal statistical education—would agree.

I should note here that this book constitutes one of the *Journal*'s first ventures in book publishing. We hope it meets the standards of quality we have always tried to maintain for the *Journal*, and that it will find favor with a broad cross-section of physicians and students.

Arnold S. Relman, M.D.
Editor, *New England Journal of Medicine*

Acknowledgments

Many people have contributed to the completion of this third edition of *Medical Uses of Statistics.* First is Fred Mosteller, who developed the vision for the first edition and extended it in the second edition. Fred worked on the present update as long as he could, and then suggested that Dave Hoaglin take his place. He was, as usual, exactly right in his assessment of who could work well with whom. We are pleased to dedicate this edition to Fred.

Jeff Drazen first suggested that Fred Mosteller and John Bailar prepare a third edition, and Jeff has been a constant source of encouragement and support through the entire process, including reading and commenting on each chapter as it reached its final stages.

Doris Peter also had a critical role; as facilitator in the later years of writing, she kept us moving ahead even when moving was difficult. Doris had an invaluable role in managing the many versions of each manuscript chapter, and in seeing those manuscripts turned into print. Without Fred, Jeff, and Doris this book would not exist.

Joe Elia provided important support and advice as this edition was being blocked out. Elizabeth Platt copy-edited the entire book. Kent Anderson, at the *New England Journal of Medicine,* and Steve Quigley, at John Wiley & Sons, worked out the details of what was necessarily a difficult and complicated sharing of responsibilities for the completion and publication of the product.

We thank John D. Emerson and Kay Larholt for timely advice.

We are grateful to all of the contributors for their hard work, dedication, and patience in writing with a level and style that were unfamiliar to almost all of them. And we are grateful to readers of the first and second editions who told us about additions and other changes that they would like to see in a future edition. We hope that readers of the present volume will follow their example.

Origins of Chapters

Chapter 1. Substantially revised, expanded, and updated for the second edition from an article originally published in the *New England Journal of Medicine* (1985; 312:890–7); slightly revised for the third edition.

Chapter 2. Based on an original publication in the Carolina Environmental Essay Series (1988; No. 9), Institute for Environmental Studies, University of North Carolina at Chapel Hill, with some new examples for this edition. Printed with permission of the publisher.

Chapter 3. Updated from the original article published in the *New England Journal of Medicine* (1983; 309:709–13) and from the second edition.*

Chapter 4. This article was written for this edition of this book. It replaces an article in the second edition.*

Chapter 5. Updated from the original article published in the *New England Journal of Medicine* (1984; 310:24–31) and from the second edition.

Chapter 6. Updated from the original article published in the *New England Journal of Medicine* (1984; 311:705–10).

Chapter 7. This article was written for this edition of this book.*

Chapter 8. This article was written for the second edition of this book and updated for this edition. The version prepared for the first edition was based heavily on material in Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. What are *P* values? In: *Biostatistics in Clinical Medicine.* New York: Macmillan Publishing Co., Inc. 1983:160–76. Printed with permission of the publisher.

Chapter 9. This article was written for this edition of this book.*

Chapter 10. This article was written for this edition of this book. It replaces an article in the second edition.*

Chapter 11. This article was written for the second edition of this book and updated and extended for this edition.

Chapter 12. This article was written for this edition of this book. It replaces an article in the second edition.*

Chapter 13. Updated and shortened from the original article published in the *New England Journal of Medicine* (1984; 311:442–8).

Chapter 14. This article is substantially revised and updated from the second edition. The original article, slightly modified for the second edition, appeared in the *Annals of Internal Medicine* (1988; 108:266–73). Printed with permission of the publisher.*

Chapter 15. Slightly revised from the original published in the *New England Journal of Medicine* (2007; 357:2189–94).

Chapter 16. This article was written for the first edition of this book, updated for the second edition, and substantially revised and updated for this edition.

Chapter 17. This article was written for this edition of this book. It replaces an article in the second edition.*

Chapter 18. This article was written for this edition of this book.*

Chapter 19. This article was written for this edition of this book.*

Chapter 20. This article was written for this edition of this book.*

Chapter 21. This article was written for this edition of this book.*

**Indicates a chapter new to this edition or completely rewritten for this edition.*

Introduction

Statistics is increasingly important to practitioners of medicine and other medical sciences, including biomedical research investigators, but changes are so rapid that their knowledge of statistical concepts, methods, and techniques may be out of date within a few years. As in the first two editions, we focus on the critical ideas, not on the mechanics. This is largely a book for the readers, not the doers, of statistics, though the latter might profit from knowing more about the nature of the procedures they use. No prior statistical knowledge is assumed. Accordingly, there are few formulas of any kind, and fewer computing formulas. Our hope is that practitioners and students of medicine and other health fields will find here the resources they need to understand the statistical methods that they encounter in the *Journal* and elsewhere in the medical literature.

Changes in the medical uses of statistics are indeed marked. Agarwal, Colditz, and Emerson show how the use of statistical methods and concepts in the *Journal* has changed from 1978–1979, to 1989, and now to 2004. They report (in Chapter 3) that a reader with no statistical knowledge beyond such simple descriptive measures as means, percentages, and variances could fully understand 27% of *Journal* articles in 1978–1979, but only 12% in 2004. Further, the kinds of statistical knowledge needed have changed markedly. Now, 66% of *Journal* papers require some knowledge of survival analysis, compared to 11% in 1978–1979. Similarly, the proportion requiring some knowledge of epidemiologic methods has increased to 53%, from only 9%. Uses of contingency tables and statistical power calculations have also seen major increases. Other methods have decreased in frequency of use, *t*-tests and Pearson correlation coefficients among them. A substantially larger proportion of papers use more than one statistical method.

Thus, the needs of readers have changed with time. The 1989 survey led to some changes in the content of the second edition of this book (1992), but the shift in *Journal* content since then requires much more substantial changes in coverage. We have replaced a chapter on clinical trials and added a second chapter, added two on statistical methods in epidemiology, and added two on statistics in genetics. Other new or replacement chapters discuss linear regression, categorical data analysis, meta-analysis, subgroup analysis, and risk analysis. We have kept a few chapters from the first and second editions because their messages are current, but the chapters on statistical thinking, statistical content of the *Journal*, cross-over designs, survival analysis, guidelines for reporting research results, and

writing about numbers have been extensively updated, and a chapter on ordered categories also has been updated and shortened. Overall, more than two-thirds of the content is new; only three chapters are substantially unchanged.

This book is meant to provide self-instruction in basic aspects of statistics as used in medicine and other health-related fields, as well as to serve as a textbook for readers who are full-time students or taking continuing education courses. With few exceptions, we stress the concepts underlying statistics rather than its more technical how-to-do-it aspects. Most of our examples come from the pages of the *New England Journal of Medicine.* We deal with both the results of investigation and the presentation of results.

Although readers can find review and didactic papers on specific statistical methods in textbooks or journals, they may not always know when or how their knowledge is incomplete or out of date, and they may have nowhere to turn for overviews of the field. This book surveys statistical applications now used in clinical research and illustrates good and poor uses of methods.

Although each chapter stands alone and can be read as a separate work, they make up five broad sections. Section I opens with a chapter (Statistical Concepts Fundamental to Investigations) on the larger concepts of statistics. That chapter surveys some of the ideas that are central to statistical methods and techniques—ideas that guide all statistical work. These broad concepts are important even when no numbers appear in a research article: Users of statistical methods should not think of numerical techniques (such as estimation or special methods of testing hypotheses) as the main ideas in statistics, while leaving the big ideas unrecognized and neglected. Chapter 2 (Some Uses of Statistical Thinking) extends the concepts in the first chapter, and illustrates with four examples how the practicalities of real life often make the uncertainty associated with statistical inferences much larger than the usual formulas for confidence limits would indicate. Such challenges arise from the need for sound data in statistical analysis, errors in critical assumptions, and uncertainty about generalizing results in complicated situations, such as moving from data acquired from animal experiments to future human experience. The chapter includes an illustration of how a complex problem can be attacked as a sequence of somewhat simpler problems. The next chapter (Use of Statistical Analysis in the *Journal*), the third in a series, tells how often various statistical procedures were used in one volume of the *Journal* and what a reader should know to understand journal reports; this chapter on frequency of use offers practical guidance to persons planning a program of study, whether they are instructors developing courses or interested readers pursuing their own education.

Section II deals with a major statistical area—the design of investigations in the medical sciences. Chapter 4 focuses on randomized trials, which have come to dominate much medical research; it discusses issues of specifying

the question, choosing the method for assigning subjects to groups, appraising the choice of outcomes, weighing the statistical power of the study, and recognizing a need to end a study early. An understanding of these matters is important to readers whether the topic is treatment, prevention, or earlier and more accurate diagnosis. Chapter 5, on crossover and self-controlled designs, deals with two related, powerful, and often under-used tools of investigation. More-detailed comment on simple reporting of experience with a series of cases is then offered in Chapter 6 (The Series of Consecutive Cases), including some discussion of the difficulties in interpreting series of cases and of precautions that can be taken to improve their strength. Chapter 7 first illustrates the extent to which the concepts and methods of epidemiology have penetrated a broad range of areas of clinical interest, then presents and discusses some questions that the reader as well as the author should consider in any medical study of human subjects. No reader can really understand the current medical literature without a good grasp of these matters.

Although an investigation must start with a study design, analysis becomes the focus after the data are in. Section III describes some central topics in data analysis. Chapter 8 (*p*-values) discusses the meaning of *p*-values, the usual way of stating the results of tests of significance, which are widely used but often misunderstood. The chapter explains the assumptions that underlie *p*-values, which have a straightforward meaning only in the presence of likely alternative hypotheses. It is most important to understand the strengths of *p*-values in terms of achieving objectivity, as well as their weaknesses for decisions or policy. Therefore, this chapter deals with both uses and misuses. Section III then turns to five specific categories of methods. Four of these deal with major types of statistical analysis in the medical sciences—linear regression (Chapter 10), survival analysis (Chapter 11), categorical data (Chapter 12), and ordered categories (Chapter 13). This section also includes further discussion of some issues in the analysis and interpretation of randomized trials (Chapter 9, which extends the discussion in Chapter 4).

The increased use of survival analysis in the clinical literature has caused us to extend the discussion of failure-time data in Chapter 11. Survival analyses must ordinarily account for the fact that not all subjects in an investigation will have experienced some key event, such as death or stroke, by the time the analysis must be made. Competing risks are explained, as are the widely used Kaplan-Meier method of estimating survival distributions and the Cox proportional-hazards model.

Contingency tables are widely used to describe patients under study and to analyze the consequences of treatment. Thus, Chapter 12 (Categorical Data) explains notions related to the 2×2 contingency table, including odds ratios, Fisher's exact test, and the paradoxes that arise when tables are collapsed. One

common generalization brings together 2×2 tables from several strata. The much-used technique of logistic regression extends the ideas of regression to situations where the outcome variable is dichotomous (0 or 1).

The chapters in this section make clear that investigators must have in mind specific questions about a set of data before they can make a rational choice of analytic methods, and that readers need to know what the investigators were after and how their goal shaped the design and analysis of a study—and what can or cannot be learned from it.

Once an investigation has been executed, the results must be conveyed Readers and investigators may find the help they need in Chapters 14, 15, and 16 in Section IV on communicating results. When faced with the masses of numbers produced by any large quantitative study, one must consider what parts of the background and results to present. Chapter 14 (Guidelines for Statistical Reporting) gives the investigator some general ideas about what to offer readers and what to keep in one's notebooks. The chapter expands on the brief statistical guidelines given as the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, published and periodically updated by the International Committee of Medical Journal Editors, and comments on some other guidelines. It gives advice about 17 specific issues that frequently arise in preparing a clinical paper containing numerical data. Chapter 15 discusses the interpretation of results seen for subgroups of a study population, which raises a vexing issue commonly known as "multiple comparisons," a matter that arises in several other chapters. The apparently simple act of writing about numbers (Chapter 16) can be much improved by understanding how to simplify, condense, and present quantitative data in text, tables, or figures. This chapter describes some common but easily avoided perils to those whose experience is primarily in working with words rather than numbers. It offers some conventions and rules about reporting numerical data.

Section V deals with five more-specialized topics. Reviewers of the literature assemble information about a particular topic from many papers. This assembly often goes beyond narrative review of the literature to a more-formal integration of quantitative information from different reports, often called meta-analysis. Chapter 17 (Combining Results) describes the various features of the research synthesis carried out by meta-analysts, illustrates the variety of methods used, and explains what a reader should be looking for in appraising a meta-analysis. Chapter 18 extends the discussion in Chapter 7 with diverse examples of regression methods applied to epidemiologic data. Chapters 19 and 20 take up a new topic, the statistical analysis of genetic data, including the investigation of hypotheses about genetic influences on human health and identifying specific genes that contribute to disease risk by genetic association studies. Chapter 21 surveys a field important to clinicians, assessing risks of various kinds to their patients.