Chapman & Hall/CRC Biostatistics Series

COMPUTATIONAL METHODS IN BIOMEDICAL RESEARCH

EDITED BY
Ravindra Khattree
Dayanand N. Naik

Chapman & Hall/CRC
Taylor & Francis Group

Computational Methods in Biomedical Research

Chapman & Hall/CRC Biostatistics Series

Editor-in-Chief

Shein-Chung Chow, Ph.D.
Professor
Department of Biostatistics and Bioinformatics
Duke University School of Medicine
Durham, North Carolina, U.S.A.

Series Editors

Byron Jones
Senior Director
Statistical Research and Consulting Centre
(IPC 193)
Pfizer Global Research and Development
Sandwich, Kent, UK

Jen-pei Liu
Professor
Division of Biometry
Department of Agronomy
National Taiwan University
Taipei, Taiwan

Karl E. Peace
Director, Karl E. Peace Center for Biostatistics
Professor of Biostatistics
Georgia Cancer Coalition Distinguished Cancer Scholar
Georgia Southern University, Statesboro, GA

Chapman & Hall/CRC Biostatistics Series

Published Titles

1. *Design and Analysis of Animal Studies in Pharmaceutical Development,* Shein-Chung Chow and Jen-pei Liu
2. *Basic Statistics and Pharmaceutical Statistical Applications,* James E. De Muth
3. *Design and Analysis of Bioavailability and Bioequivalence Studies, Second Edition, Revised and Expanded,* Shein-Chung Chow and Jen-pei Liu
4. *Meta-Analysis in Medicine and Health Policy,* Dalene K. Stangl and Donald A. Berry
5. *Generalized Linear Models: A Bayesian Perspective,* Dipak K. Dey, Sujit K. Ghosh, and Bani K. Mallick
6. *Difference Equations with Public Health Applications,* Lemuel A. Moyé and Asha Seth Kapadia
7. *Medical Biostatistics,* Abhaya Indrayan and Sanjeev B. Sarmukaddam
8. *Statistical Methods for Clinical Trials,* Mark X. Norleans
9. *Causal Analysis in Biomedicine and Epidemiology: Based on Minimal Sufficient Causation,* Mikel Aickin
10. *Statistics in Drug Research: Methodologies and Recent Developments,* Shein-Chung Chow and Jun Shao
11. *Sample Size Calculations in Clinical Research,* Shein-Chung Chow, Jun Shao, and Hansheng Wang
12. *Applied Statistical Design for the Researcher,* Daryl S. Paulson
13. *Advances in Clinical Trial Biostatistics,* Nancy L. Geller
14. *Statistics in the Pharmaceutical Industry, Third Edition,* Ralph Buncher and Jia-Yeong Tsay
15. *DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments,* David B. Allsion, Grier P. Page, T. Mark Beasley, and Jode W. Edwards
16. *Basic Statistics and Pharmaceutical Statistical Applications, Second Edition,* James E. De Muth
17. *Adaptive Design Methods in Clinical Trials,* Shein-Chung Chow and Mark Chang
18. *Handbook of Regression and Modeling: Applications for the Clinical and Pharmaceutical Industries,* Daryl S. Paulson
19. *Statistical Design and Analysis of Stability Studies,* Shein-Chung Chow
20. *Sample Size Calculations in Clinical Research, Second Edition,* Shein-Chung Chow, Jun Shao, and Hansheng Wang
21. *Elementary Bayesian Biostatistics,* Lemuel A. Moyé
22. *Adaptive Design Theory and Implementation Using SAS and R,* Mark Chang
23. *Computational Pharmacokinetics,* Anders Kallen

Chapman & Hall/CRC Biostatistics Series

Computational Methods in Biomedical Research

Edited by

Ravindra Khattree
Oakland University
Rochester, Michigan, U.S.A.

Dayanand N. Naik
Old Dominion University
Norfolk, Virginia, U.S.A.

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an **informa** business

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-58488-577-1 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Computational methods in biomedical research / editors, Ravindra Khattree and Dayanand Naik.
p. ; cm. -- (Biostatistics series ; 24)
"A CRC title."
Includes bibliographical references and index.
ISBN 978-1-58488-577-1 (alk. paper)
1. Medicine--Research--Data processing. 2. Biology--Research--Data processing. 3. Medicine--Research--Statistical methods. 4. Biology--Research--Statistical methods. 5. Computational biology. I. Khattree, Ravindra. II. Naik, Dayanand N. III. Series: Chapman & Hall/CRC biostatistics series ; 24.
[DNLM: 1. Computational Biology--methods. 2. Biomedical Research--methods. 3. Data Interpretation, Statistical. QU 26.5 C73756 2008]

R853.D37C63 2008
610.285--dc22 2007029936

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Series Introduction

The primary objectives of the Biostatistics Book Series are to provide useful reference books for researchers and scientists in academia, industry, and government, and also to offer textbooks for undergraduate and graduate courses in the area of biostatistics. This book series will provide comprehensive and unified presentations of statistical designs and analyses of important applications in biostatistics, such as those in biopharmaceuticals. A well-balanced summary will be given of current and recently developed statistical methods and interpretations for both statisticians and researchers or scientists with minimal statistical knowledge who are engaged in the field of applied biostatistics. The series is committed to providing easy-to-understand, state-of-the-art references and textbooks. In each volume, statistical concepts and methodologies will be illustrated through real-world examples.

In the last decade, it was recognized that increased spending on biomedical research does not reflect an increase in the success rate of pharmaceutical development. On March 16, 2004, the FDA released a report addressing the recent slowdown in innovative medical therapies submitted to the FDA for approval, "Innovation/Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products." The report describes the urgent need to modernize the medical product development process—the critical path—to make product development more predictable and less costly. Two years later, the FDA released a Critical Path Opportunities List that outlines 76 initial projects (under six broad topic areas) to bridge the gap between the quick pace of new biomedical discoveries and the slower pace at which those discoveries are currently developed into therapies. Among the six broad topic areas, better evaluation tool (development of biomarker), streamlining clinical trial (the use of adaptive design methods), and harnessing bioinformatics (the use of computational biology) are considered the top three challenges for increasing the probability of success in pharmaceutical research and development.

This volume provides useful approaches for implementation of target clinical trials in pharmaceutical research and development. It covers statistical methods for various computational topics such as biomarker development, sequential monitoring, proportional hazard mixed-effects models, and Bayesian approach in pharmaceutical research and development. It would be beneficial to biostatisticians, medical researchers, pharmaceutical

scientists, and reviewers in regulatory agencies who are engaged in the areas of pharmaceutical research and development.

Shein-Chung Chow

Preface

This edited volume is a collection of chapters covering some of the important computational topics with special reference to biomedical applications. Rapid advances in ever-changing biomedical research and methodological statistical developments that must support these advances make it imperative that from time to time a cohesive account of new computational schemes is made available for users to implement these methodologies in the particular biomedical context or problem. The present volume is an attempt to fill this need.

Realizing the vastness of the area itself, there is no pretension to be exhaustive in terms of the general field or even in terms of a topic represented by a chapter within this field; such a task, while also requiring hundreds of collaborators, would require a collection of several volumes of similar size. Hence the selection made here represents our personal view of what the most important topics are, in terms of their applicability and potential in the near future. With this in mind, the chapters are arranged accordingly, with the works of immediate applicability appearing first. These are followed by more theoretical advances and computational schemes that are yet to be developed in satisfactory forms for general applications.

Work of this magnitude could not have been accomplished without the help of many people. We wish to thank our referees for painstakingly going through the chapters as a gesture of academic goodwill. Theresa Del Forn of Taylor & Francis Group, was most helpful and patient with our repeatedly broken promises of meeting the next deadline. Our families have provided their sincere support during this project and we appreciate their understanding as well.

Ravindra Khattree, Rochester, Michigan
Dayanand N. Naik, Norfolk, Virginia

Editors

Ravindra Khattree, professor of statistics at Oakland University, Rochester, Michigan, received his initial graduate training at the Indian Statistical Institute. He received his PhD from the University of Pittsburgh in 1985. He is the author or coauthor of numerous research articles on theoretical and applied statistics in various national and international journals. His research interests include multivariate analysis, experimental designs, quality control, repeated measures, and statistical inference. In addition to teaching graduate and undergraduate courses, Dr. Khattree regularly consults with industry, hospitals, and academic researchers on various applied statistics problems. He is a chief editor of the *Journal of Statistics and Applications*, editor of *InterStat*, an online statistics journal, and an associate editor of the *Journal of Statistical Theory and Practice*. For many years, he also served as an associate editor for the *Communications in Statistics*. He is a Fellow of the American Statistical Association, an elected member of the International Statistical Institute, and a winner of the Young Statistician Award from International Indian Statistical Association. Dr. Khattree is a coauthor of two books, both with Dr. D. N. Naik, titled *Applied Multivariate Statistics with SAS Software (Second Edition)* and *Multivariate Data Reduction and Discrimination with SAS Software*, both copublished by SAS Press/Wiley. He has also coedited, with Dr. C. R. Rao, the *Handbook of Statistics 22: Statistics in Industry*, published by North Holland.

Dayanand N. Naik is professor of statistics at Old Dominion University, Norfolk, Virginia. He received his MS degree in statistics from Karnatak University, Dharwad, India, and PhD in statistics from the University of Pittsburgh in 1985. He has published extensively in several well-respected journals and advised numerous students for their PhD in statistics. His research interests include multivariate analysis, linear models, quality control, regression diagnostics, repeated measures, and growth curve models. Dr. Naik is an editor of *InterStat*, a statistics journal on the Internet, and an associate editor of *Communications in Statistics*. Dr. Naik is also actively involved in statistical consulting and collaborative research. He is an elected member of International Statistical Institute and is very active in American Statistical Association activities. Dr. Naik is coauthor of two books, both with Dr. Khattree, titled *Applied Multivariate Statistics with SAS Software (Second Edition)* and *Multivariate Data Reduction and Discrimination with SAS Software*, both copublished by SAS Press/Wiley.

Contributors

Mousumi Banerjee Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan

N. Rao Chaganty Department of Mathematics and Statistics, Old Dominion University, Norfolk, Virginia

Kathleen A. Cronin Statistical Research and Applications Branch, National Cancer Institute, Bethesda, Maryland

Somnath Datta Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, Kentucky

Susmita Datta Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, Kentucky

Michael Donohue Division of Biostatistics and Bioinformatics, Department of Family and Preventive Medicine, University of California at San Diego, La Jolla, California

Eric J. Feuer Statistical Research and Applications Branch, National Cancer Institute, Bethesda, Maryland

Joseph C. Gardiner Division of Biostatistics, Department of Epidemiology, Michigan State University, East Lansing, Michigan

Kaushik Ghosh Department of Mathematical Sciences, University of Nevada at Las Vegas, Las Vegas, Nevada

Pulak Ghosh Department of Mathematics and Statistics, Georgia State University, Atlanta, Georgia

Ahmedin Jemal Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, Georgia

Ravindra Khattree Department of Mathematics and Statistics, Oakland University, Rochester, Michigan

Lin Liu Division of Biostatistics, Department of Epidemiology, Michigan State University, East Lansing, Michigan

Lin Liu Department of Family and Preventive Medicine, University of California at San Diego, La Jolla, California

Hedibert F. Lopes Graduate School of Business, University of Chicago, Chicago, Illinois

Zhehui Luo Division of Biostatistics, Department of Epidemiology, Michigan State University, East Lansing, Michigan

Deepak Mav Constella Group, Inc., Durham, North Carolina

Jarosław Meller Division of Biomedical Informatics, Cincinnati Children's Hospital Research Foundation, Cincinnati, Ohio; Department of Informatics, Nicholas Copernicus University, Torun, Poland; and Department of Environmental Health, University of Cincinnati, Ohio

Peter Müller Department of Biostatistics, The University of Texas, M. D. Anderson Cancer Center, Houston, Texas

Dayanand N. Naik Department of Mathematics and Statistics, Old Dominion University, Norfolk, Virginia

Anne-Michelle Noone Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan

Rudolph S. Parrish Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, Kentucky

Nalini Ravishanker Department of Statistics, University of Connecticut, Storrs, Connecticut

William F. Rosenberger Department of Statistics, The Volgenau School of Information Technology and Engineering, George Mason University, Fairfax, Virginia

Anuradha Roy Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, Texas

Caryn M. Thompson Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, Kentucky

Ram C. Tiwari Statistical Research and Applications Branch, National Cancer Institute, Bethesda, Maryland

Florin Vaida Division of Biostatistics and Bioinformatics, Department of Family and Preventive Medicine, University of California at San Diego, La Jolla, California

Michael Wagner Division of Biomedical Informatics, Cincinnati Children's Hospital Research Foundation, Cincinnati, Ohio

Ronghui Xu Division of Biostatistics and Bioinformatics, Department of Family and Preventive Medicine, and Department of Mathematics, University of California at San Diego, La Jolla, California

Yanqiong Zhang Merck & Company, Rahway, New Jersey

1

Microarray Data Analysis

Susmita Datta, Somnath Datta, Rudolph S. Parrish, and Caryn M. Thompson

CONTENTS