ICSA Book Series in Statistics
Series Editors: Jiahua Chen · Ding-Geng (Din) Chen

Yichuan Zhao
Ding-Geng Chen *Editors*

New Frontiers of Biostatistics and Bioinformatics

Springer

ICSA Book Series in Statistics

The ICSA Book Series in Statistics showcases research from the International Chinese Statistical Association that has an international reach. It publishes books in statistical theory, applications, and statistical education. All books are associated with the ICSA or are authored by invited contributors. Books may be monographs, edited volumes, textbooks and proceedings.

More information about this series at http://www.springer.com/series/13402

Yichuan Zhao • Ding-Geng Chen
Editors

New Frontiers of Biostatistics and Bioinformatics

Springer

Editors
Yichuan Zhao
Department of Mathematics and Statistics
Georgia State University
Atlanta, GA, USA

Ding-Geng Chen
Department of Biostatistics
Gillings School of Global Public Health
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

Department of Statistics
University of Pretoria
Pretoria, South Africa

ISSN 2199-0980 ISSN 2199-0999 (electronic)
ICSA Book Series in Statistics
ISBN 978-3-319-99388-1 ISBN 978-3-319-99389-8 (eBook)
https://doi.org/10.1007/978-3-319-99389-8

Library of Congress Control Number: 2018962164

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is mainly comprised of excellent presentations delivered in the 5th Workshop on Biostatistics and Bioinformatics held in Atlanta on May 5–7, 2017. Biostatistics and bioinformatics have been playing a key role in statistics and other scientific research fields in recent years. The aim of the 5th Workshop on Biostatistics and Bioinformatics was to stimulate research, foster interaction among researchers in field, and offer opportunities for learning and facilitating research collaborations in the era of big data. From this successful workshop, the two editors selected excellent presentations for this book. All the 22 chapters are peer-reviewed and revised multiple times before the final acceptance. This book provides the most recent advances in the field, presenting new methods and case applications at the frontiers of biostatistics and bioinformatics research and interdisciplinary areas. This timely book makes invaluable contributions to biostatistics and bioinformatics and offers insights for researchers, students, and industry practitioners.

The 22 chapters are organized into 5 parts. Part I includes five chapters that present a review of the theoretical framework in biostatistics. Part II consists of four chapters on wavelet-based approach for complex data. Part III is composed of six chapters that present clinical trials and statistical modeling. Part IV outlines high-dimensional gene expression data analysis. Part V consists of four chapters on survival analysis. We organize the chapters as self-contained units, and the references of the chapter are at the end of the each chapter so that readers can refer to the cited sources easily. To better understand the proposed procedures in the book, the readers can readily request the data sets and computer programs from the two editors. Therefore, the readers can apply these new statistical methods of the book for their own research.

Part I: Review and Theoretical Framework in Biostatistics (Chaps. 1–4)

The chapter "Optimal Weighted Wilcoxon–Mann–Whitney Test for Prioritized Outcomes" reviews concepts of prioritized outcomes in a two-group randomized clinical trial of multiple outcomes, where mortality affects the assessment of the other follow-up outcomes. In this chapter, Matsouaka, Singhal, and Betensky develop a weighted Wilcoxon–Mann–Whitney test procedure to analyze the data and determine the optimal weights that maximize its power. The authors obtain the analytical power formula for the test statistic and compare its results with those obtained via simulation studies using a range of treatment effects on the outcomes.

In the chapter "A Selective Overview of Semiparametric Mixture of Regression Models," Xiang and Yao conduct a systematic overview of new semiparametric mixture of regression models, which have been popularly used in many applications. Recent advances and open questions are also discussed.

In the chapter "Rank-Based Empirical Likelihood for Regression Models with Responses Missing at Random," Bindele and Zhao consider a general regression model with responses missing at random. From an imputed rank-based objective function, the authors derive a rank-based estimator, and its asymptotic distribution is established. An empirical likelihood approach is proposed based on the rank-based objective function, from which its asymptotic distribution is established.

In the chapter "Bayesian Nonparametric Spatially Smoothed Density Estimation," Hanson, Zhou, and de Carvalho develop a Bayesian nonparametric density estimator, which changes smoothly in space. The estimator is built using the predictive rule from a marginalized Polya tree so that observations are spatially weighted by their distance from the location of interest. The authors propose a simple refinement to accommodate arbitrarily censored data and develop a test for whether the density is spatially varying.

Part II: Wavelet-Based Approach for Complex Data (Chaps. 5–8)

The chapter "Mammogram Diagnostics Using Robust Wavelet-Based Estimator of Hurst Exponent" presents the robust estimation of Hurst exponent in two-dimensional images based on non-decimated wavelet transforms. The properties of the proposed estimators are studied both theoretically and numerically. In this chapter, Feng, Mei, and Vidakovic show how to apply proposed methods to digitized mammogram images, estimate Hurst exponent, and then use it as a discriminatory descriptor to classify mammograms to benign and malignant.

In the chapter "Wavelet-Based Profile Monitoring Using Order-Thresholding Recursive CUSUM Schemes," Zhang, Mei, and Shi propose a novel wavelet-

based profile monitoring procedure, which is based on the order-thresholding transformation of recursive CUSUM statistics of multiple wavelet coefficients. The authors carry out extensive simulation studies and a case study of tonnage profile data, which show that proposed procedure is efficient for detecting the unknown local changes on the profile.

In the chapter "Estimating the Confidence Interval of Evolutionary Stochastic Process Mean from Wavelet-Based Bootstrapping," de Medeiros and de Souza propose to estimate the uncertainty for the evolutionary mean of a stochastic process based on bootstrapping of wavelet coefficients. By discrete wavelet transform, the authors apply bootstrap to estimate the confidence interval of the autocorrelation for a time series. Moreover, these methods with few modifications are implemented.

In the chapter "A New Wavelet-Based Approach for Mass Spectrometry Data Classification," Cohen, Messaoudi, and Badir propose a statistical methodology of a reliable diagnosis for classifying mass spectrometry data with a type of cancer. The authors go over wavelets, principal component analysis, and support vector machines, and perform a study on low-mass SELDI spectra from patients with breast cancer and from normal controls. The performance is evaluated with a k-fold cross validation technique and simulation study. The performance of the proposed method is excellent with an accurate classification of mass spectrometry.

Part III: Clinical Trials and Statistical Modeling (Chaps. 9–14)

In the chapter "Statistical Power and Bayesian Assurance in Clinical Trial Design," Chen and Chen propose a Bayesian assurance as an alternative to the conventional statistical power to incorporate the uncertainties of this observed treatment effect. In this chapter, the authors review the transition from conventional statistical power to Bayesian assurance and discuss the computations of Bayesian assurance using a Monte Carlo simulation-based method.

The chapter "Equivalence Tests in Subgroup Analyses" proposes that the consistency of the treatment effect in two subgroups should be assessed using an equivalence test called *consistency* test. In this chapter, Ring, Scharpenberg, Grill, Schall, and Brannath present tests for both quantitative and binary outcome variables and review the basic properties of these consistency tests using simulation studies. The authors also indicate that equivalence tests can be used both to assess the consistency of treatment effects across subgroups and to detect medically relevant heterogeneity in treatment effects across subgroups.

In the chapter "Predicting Confidence Interval for the Proportion at the Time of Study Planning in Small Clinical Trials," Yu and Vexler discuss "future" confidence interval prediction with binomial outcomes for small clinical trials and sample size calculation, where the "future" confidence interval emphasizes the confidence interval as a function of a random sample that is not observed at the planning

stage of a study. The authors propose three probabilistic approaches to future confidence interval prediction when the sample size is small. The approach based on the expectation of the boundaries has the most desirable properties and is easy to implement.

The chapter "Importance of Adjusting for Multi-Stage Design When Analyzing Data from Complex Surveys" illustrates possible discrepancies in point estimates and standard errors using 2014–2015 TUS data. In this chapter, Ha and Soulakova show the importance of using the guidelines when analyzing complex surveys. The authors discuss three methods: method I ignores any weighting, method II incorporates the main weight only, and method III utilizes the main weight and balanced repeated replications with specified replicate weights.

In the chapter "Analysis of the High School Longitudinal Study to Evaluate Associations Among Mathematics Achievement, Mentorship and Student Participation in STEM Programs," Murillo, Tiwari, and Affuso analyze a subsample of the High School Longitudinal Study (2009–2013) dataset (HSLS:09). Regression models are applied to evaluate mathematics achievement and student enrollment in STEM major/careers based on their individual participation. Differences based on sex, race/ethnicity, and socioeconomic status are assessed.

The chapter "Statistical Modeling for the Heart Disease Diagnosis via Multiple Imputation" addresses a common challenge of missing data during statistical analysis of clinic data. Missing data causes severe problems in statistical analysis and leads to invalid conclusions. Multiple imputation is a useful strategy for handling missing data. In the chapter, Li and Zhao apply the multiple imputation to a public accessible heart disease dataset, which has a high missing rate, and build a prediction model for the heart disease diagnosis.

Part IV: High-Dimensional Gene Expression Data Analysis (Chaps. 15–18)

In the chapter "Learning Gene Regulatory Networks with High-Dimensional Heterogeneous Data," Jia and Liang propose to model the heterogeneous data using a mixture Gaussian graphical model and apply the imputation-consistency algorithm to estimate the parameters of the mixture model and cluster the samples to different subgroups. The proposed method is compared with an existing method for learning mixture Gaussian graphical models as well as a few other methods for homogeneous data, such as graphical Lasso, etc.

The chapter "Performance Evaluation of Normalization Approaches for Metagenomic Compositional Data on Differential Abundance Analysis" assesses normalization methods for metagenomic sequence data. In this chapter, Du, An, and Fang further study the impact of normalization on subsequent differential abundance analysis. The authors suggest the selection of a normalization method for metagenomic compositional data should be made on a case-by-case basis.

The chapter "Identification of Pathway-Modulating Genes Using the Biomedical Literature Mining" centers on an effective use of biomedical literature for the identification of the relationships among genes. A Bayesian hierarchical model was proposed, which allows to identify indirect relationship between genes by linking them using the gene ontology terms. In this chapter, Yu, Nam, Couch, Lawson, and Chung illustrate this method using the web interface GAIL. It provides the PubMed literature mining results, along with the R package by the Bayesian hierarchical model.

The chapter "Discriminant Analysis and Normalization Methods for Next-Generation Sequencing Data" studies discriminating and normalization methods for gene expression analysis with the development of high-throughput techniques. A number of new discriminant analysis methods have been proposed to discriminate next-generation sequencing data. In this chapter, Zhou, Wang, Zhao, and Tong introduce three methods including the Poisson linear discriminant analysis, the zero-inflated Poisson logistic discriminant analysis, and the negative binomial linear discriminant analysis and further introduce several normalization methods for processing next-generation sequencing data.

Part V: Survival Analysis (Chaps. 19–22)

In the chapter "On the Landmark Survival Model for Dynamic Prediction of Event Occurrence Using Longitudinal Data," Zhu, Li, and Huang demonstrate that a joint distribution of longitudinal and survival data exists that satisfy the modeling assumptions without additional restrictions. In addition, the authors propose an algorithm to generate data from this joint distribution and generalize the results to the more flexible landmark linear transformation models, which include the landmark Cox model.

In the chapter "Nonparametric Estimation of a Cumulative Hazard Function with Right Truncated Data," Zhang, Jiang, Zhao, and Akcin develop the nonparametric inference for the forward-time hazard. The authors study a family of weighted tests for comparing the hazard function between two independent samples. The authors analyze the data set about AIDS incubation time to illustrate the proposed procedures.

In the chapter "Empirical Study on High-Dimensional Variable Selection and Prediction Under Competing Risks," Hou and Xu consider competing risk analysis and explore statistical properties in the presence of high-dimensional predictors. The authors study the accuracy of prediction and variable selection of existing statistical learning methods using extensive simulation studies, including different approaches to choosing penalty parameters in each method.

In the chapter "Nonparametric Estimation of a Cumulative Hazard Function with Right Truncated Data," Akcin, Zhang, and Zhao study the nonparametric inference for the hazard rate function with right truncated data. Kernel smoothing techniques

are used to get smoothed estimates of hazard rates. Three commonly used kernels, uniform, Epanechnikov, and biweight kernels are applied on the AIDS data to illustrate the proposed methods.

We are very grateful to all of people, who have supported the creation of this book with Springer. First, we thank the authors of each chapter for their wonderful contributions. Second, our deep gratitude goes to all the reviewers for their dedicated reviews, which improved the quality of the book significantly. Third, we would like to acknowledge the leadership of the organizing committee and all the volunteers of the 5th Workshop on Biostatistics and Bioinformatics because this book would be impossible without this workshop. Last but not least, our sincere appreciations go to the professional support and great assistance of Nicholas Philipson (Springer/ICSA Book Series coordinator and editorial director, Business/Economics & Statistics), Nitza Jones-Sepulveda (associate editor) from Springer New York, and Sindhuraj Thulasingam (Project Coordinator of Books) from Springer Nature, which made this book published. We welcome readers' comments on typos, errors, and improvements about the book. If there is an exchange, please send comments and suggestions to Dr. Yichuan Zhao (email: yichuan@gsu.edu) and Dr. Ding-Geng Chen (email: dinchen@email.unc.edu).

Atlanta, GA, USA Yichuan Zhao
Chapel Hill, NC, USA Ding-Geng Chen

Contents

List of Contributors

Olivia Affuso Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA

Haci Akcin Department of Risk Management and Insurance, Georgia State University, Atlanta, GA, USA

Lingling An Department of Agricultural and Biosystems Engineering, University of Arizona, Tucson, AZ, USA

Hassan Badir National School of Applied Sciences-Tangier, ENSAT, Abdelmalek Essaadi University, Tangier, Morocco

Rebecca A. Betensky Department of Biostatistics, Harvard T.H. Chan School of Public Health, and Harvard NeuroDiscovery Center, Harvard Medical School, Boston, MA, USA

Huybrechts F. Bindele Department of Mathematics and Statistics, University of South Alabama, Mobile, AL, USA

W. Brannath Faculty of Mathematics/Computer Sciences, Competence Center for Clinical Trials Bremen, University of Bremen, Bremen, Germany

Ding-Geng Chen Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Department of Statistics, University of Pretoria, Pretoria, South Africa

Jenny K. Chen Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Dongjun Chung Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

Achraf Cohen Department of Mathematics and Statistics, University of West Florida, Pensacola, FL, USA

Daniel Couch Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

Vanda Inácio de Carvalho University of Edinburgh, Edinburgh, Scotland, UK

Aline Edlaine de Medeiros Graduate Program in Biostatistics, State University of Maringa, Maringa, Brazil

Eniuce Menezes de Souza Department of Statistics, State University of Maringa, Maringa, Brazil

Ruofei Du Biostatistics Shared Resource, University of New Mexico Comprehensive Cancer Center, Albuquerque, NM, USA

Zhide Fang Biostatistics Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA, USA

Chen Feng H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

S. Grill Faculty of Mathematics/Computer Science, Competence Center for Clinical Trials Bremen, University of Bremen, Bremen, Germany

Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

Trung Ha Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, Orlando, FL, USA

Timothy Hanson Medtronic Inc., Minneapolis, MN, USA

Jiayi Hou Altman Clinical and Translational Research Institute, University of California, San Diego, La Jolla, CA, USA

Xuelin Huang Department of Biostatistics, MD Anderson Cancer Center, Houston, TX, USA

Bochao Jia Eli Lilly and Company, Lilly Corporate Center, IN, USA

Yong Jiang MetLife Inc., Whippany, NJ, USA

Andrew Lawson Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

Lian Li Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

Liang Li Department of Biostatistics, MD Anderson Cancer Center, Houston, TX, USA

Faming Liang Department of Statistics, Purdue University, West Lafayette, IN, USA

Roland A. Matsouaka Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

Program for Comparative Effectiveness Methodology, Duke Clinical Research Institute, Duke University, Durham, NC, USA

Yajun Mei H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Chaimaa Messaoudi National School of Applied Sciences-Tangier, ENSAT, Abdelmalek Essaadi University, Tangier, Morocco

Anarina L. Murillo Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA

Jin Hyun Nam Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

Arne Ring Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa

medac GmbH, Wedel, Germany

R. Schall Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa

IQVIA Biostatistics, Bloemfontein, South Africa

M. Scharpenberg Faculty of Mathematics/Computer Sciences, Competence Center for Clinical Trials Bremen, University of Bremen, Bremen, Germany

Jianjun Shi H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Aneesh B. Singhal Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

Julia N. Soulakova Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, Orlando, FL, USA

Hemant K. Tiwari Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA

Tiejun Tong Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong

Albert Vexler Department of Biostatistics, University at Buffalo, State University of New York, Buffalo, NY, USA

Brani Vidakovic H. Milton Stewart School of Industrial and Systems Engineering and Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Junhui Wang School of Data Science, City University of Hong Kong, Kowloon, Hong Kong

Sijia Xiang School of Data Sciences, Zhejiang University of Finance & Economics, Hangzhou, China

Ronghui Xu Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA

Department of Mathematics, University of California, San Diego, La Jolla, CA, USA

Weixin Yao Department of Statistics, University of California, Riverside, CA, USA

Jihnhee Yu Department of Biostatistics, University at Buffalo, State University of New York, Buffalo, NY, USA

Zhenning Yu Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

Ruizhi Zhang H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Xu Zhang Center for Clinical and Translational Sciences, University of Texas Health Science Center, Houston, TX, USA

Yichuan Zhao Department of Mathematics and Statistics, Georgia State University, Atlanta, GA, USA

Haiming Zhou Northern Illinois University, DeKalb, IL, USA

Yan Zhou College of Mathematics and Statistics, Institute of Statistical Sciences, Shenzhen University, Shenzhen, China

Yayuan Zhu Department of Biostatistics, MD Anderson Cancer Center, Houston, TX, USA

List of Chapter Reviewers

David Benkeser Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA

Frazier Bindele Department of Mathematics and Statistics, University of South Alabama, Mobile, AL, USA

Qingpo Cai Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA

Hsin-wen Chang Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Ding-Geng Chen School of Social Work, University of North Carolina, Chapel Hill, NC, USA

Yunxiao Chen Department of Psychology and Institute for Quantitative Theory and Methods, Emory University, Atlanta, GA, USA

Eric Chicken Department of Statistics, Florida State University, Tallahassee, FL, USA

Xin Dang Department of Mathematics, University of Mississippi, University, MS, USA

Yang Feng Department of Statistics, Columbia University, New York, NY, USA

Cindy Fu Department of Mathematics and Statistics, York University, Toronto, ON, Canada

Matt Hayat Division of Epidemiology and Biostatistics, School of Public Health, Georgia State University, Atlanta, GA, USA

Zhiguang Huo Department of Biostatistics, University of Florida, Gainesville, FL, USA

Hsin-Hsiung Bill Huang Department of Statistics, University of Central Florida, Orlando, FL, USA

Linyuan Li Department of Mathematics and Statistics, University of New Hampshire, Durham, NH, USA

Pengfei Li Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada

Jason Liao Nonclinical Statistics, Merck & Co., Inc., Whitehouse Station, NJ, USA

Antonio Linero Department of Statistics, Florida State University, Tallahassee, FL, USA

Yang Liu Statistics, Programming, and Economics Branch, Division of Analysis Research and Practice Integration, National Center for Injury Prevention and Control, Atlanta, GA, USA

Xuewen Lu Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada

Sheng Luo Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, USA

Xiaoyi Min Department of Mathematics and Statistics, Georgia State University, Atlanta, GA, USA

Timothy O'Brien Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL, USA

Li-Xuan Qin Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Lixin Shen Department of Mathematics, Syracuse University, Syracuse, NY, USA

Chuck Song Division of Biostatistics, College of Public Health, Ohio State University, Columbus, OH, USA

Xin Tian Office of Biostatistics Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA

Brani Vidakovic H. Milton Stewart School of Industrial and Systems Engineering, and Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Antai Wang Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ, USA

Dongliang Wang Department of Public Health and Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY, USA

Lianming Wang Department of Statistics, University of South Carolina, Columbia, SC, USA

Weizheng Wang Department of Mathematics and Statistics, Wright State University, Dayton, OH, USA

Wei (Peter) Yang Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Lili Yu Department of Biostatistics, Georgia Southern University, Statesboro, GA, USA

Jiajia Zhang Department of Statistics, University of South Carolina, Columbia, SC, USA

Mei-Jie Zhang Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI, USA

Min Zhang Department of Statistics, Purdue University, West Lafayette, IN, USA

Xu Zhang Division of Clinical and Translational Sciences, Department of Internal Medicine, Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA

Ying-Ying Zhang Department of Statistics and Actuarial Science, College of Mathematics and Statistics, Chongqing University, Chongqing, China

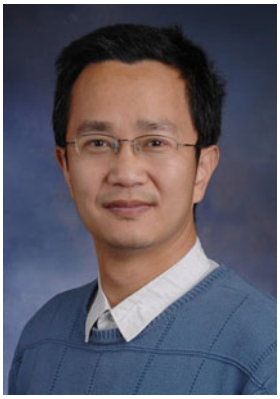
Yichuan Zhao Department of Mathematics and Statistics, Georgia State University, Atlanta, GA, USA

Yunpeng Zhao Department of Statistics, Volgenau School of Engineering, George Mason University, Fairfax, VA, USA

Hongjian Zhu Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, Houston, TX, USA

Changliang Zou Institute of Statistics, Nankai University, Tianjin, China

About the Editors



Yichuan Zhao is a professor of statistics at Georgia State University in Atlanta. He has a joint appointment as associate member of the Neuroscience Institute, and he is also an affiliated faculty member of School of Public Health at Georgia State University. His current research interest focuses on survival analysis, empirical likelihood method, nonparametric statistics, analysis of ROC curves, bioinformatics, Monte Carlo methods, and statistical modeling of fuzzy systems. He has published over 80 research articles in statistics; has co-edited two books on statistics, biostatistics, and data science; and has been invited to deliver more than 170 research talks nationally and internationally. Dr. Zhao has organized the Workshop Series on Biostatistics and Bioinformatics since its initiation in 2012. He also organized the 25th ICSA Applied Statistics Symposium in Atlanta as a chair of the organizing committee to great success. He is currently serving as editor, or on the editorial board, for several statistical journals. Dr. Zhao is an elected member of the International Statistical Institute.

(Din) Ding-Geng Chen is a fellow of the American Statistical Association and currently the Wallace H. Kuralt distinguished professor at the University of North Carolina at Chapel Hill, USA, and an extraordinary professor at the University of Pretoria, South Africa. He was a professor at the University of Rochester and the Karl E. Peace endowed eminent scholar chair in biostatistics at Georgia Southern University. He is also a senior consultant for biopharmaceuticals and government agencies with extensive expertise in clinical trial biostatistics and public health statistics with multimillion dollars' federal-funded research projects. Professor Chen has written more than 150 refereed publications and co-authored/co-edited 23 books on clinical trial methodology, meta-analysis, causal-inference, and public health statistics.

Part I
Review of Theoretical Framework in Biostatistics

Chapter 1
Optimal Weighted Wilcoxon–Mann–Whitney Test for Prioritized Outcomes

Roland A. Matsouaka, Aneesh B. Singhal, and Rebecca A. Betensky

This chapter reviews keys concepts of prioritized outcomes in a two-group randomized clinical trial of multiple outcomes, where mortality affects the assessment of the other follow-up outcomes. The main concepts related to prioritized endpoints along with the different terminologies used in the literature are discussed. Then, statistical tenets of worst-rank composite endpoints are reviewed using a combined endpoint of mortality and a continuous outcome.

We motivate the approach using a randomized clinical trial of normobaric oxygen therapy on patients who underwent an acute ischemic stroke where we combine a continuous outcome with mortality into a single composite endpoint using the worst-rank framework. We develop a weighted Wilcoxon–Mann–Whitney test statistic to analyze the data and determine the optimal weights that maximize its power. We provide the rationale for the weights and their relative importance in data analysis. In addition, we derive the analytical power formula for the test statistic. To demonstrate that the proposed power formula produces valid power estimations, we compare its results with those obtained empirically via Monte-Carlo simulations using a range of treatment effects on the outcomes. Finally, we illustrate the method using data from the clinical trial of normobaric oxygen therapy.

R. A. Matsouaka (✉)
Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA
e-mail: roland.matsouaka@duke.edu

Program for Comparative Effectiveness Methodology, Duke Clinical Research Institute, Duke University, Durham, NC, USA

A. B. Singhal
Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

R. A. Betensky
Department of Biostatistics, Harvard T.H. Chan School of Public Health and Harvard NeuroDiscovery Center, Harvard Medical School, Boston, MA, USA
e-mail: betensky@hsph.harvard.edu

Y. Zhao, D.-G. Chen (eds.), *New Frontiers of Biostatistics and Bioinformatics*,
ICSA Book Series in Statistics, https://doi.org/10.1007/978-3-319-99389-8_1

1.1 Introduction

In clinical trials of multifaceted diseases, multiple outcomes are usually evaluated to estimate and compare the effects of a new active treatment over a control treatment. Although these outcomes can be analyzed separately, they are usually combined into a single composite endpoint to take into account the complexity of the disease manifestations and capture different aspects of the treatment effects. Combining outcomes has several advantages: it increases statistical precision and efficacy, reduces considerably the number of patients needed to enroll for a given expected treatment effect or to reach a specific statistical power, circumvents the needs for multiple testing, and provides an overall assessment of the treatment effect.

One of the most commonly used methods for combining multiple outcomes is the time-to-first event. For this method, only a patient's initial event during the trial is considered in the analysis while all of the subsequent events are ignored. However, such a composite endpoint have serious practical limitations that often result in misleading interpretations and poor medical decisions which are of greater concerns. Usually, component outcomes of a composite endpoint are not equally important or clinically relevant; they do not occur at the same frequency and are not similarly impacted by the treatment. More than often, treatment effects and significant statistical analyses are driven by components of lesser importance. As such, they do not provide a more comprehensive perspective of the disease burden that is realistic, congruent with clinical judgment or aligned with the perceptions and expectations of patients and their caregivers.

This is illustrated in many cardiovascular disease trials where mortality remains the major outcome of interest which, fortunately, is often less frequent and tends to occur later in a trial (see, for instance, the relative perceived clinical severity of typical components of composite endpoints considered in recent cardiovascular trials given in Fig. 1.1). In a clinical trial of "death or heart failure hospitalization" (whichever comes first), for example, a patient may experience multiple heart failure hospitalizations and eventually die. Clearly, a patient who has a minor heart attack after 1 week of follow-up but remain event-free subsequently for several consecutive years should not be considered as having a worse outcome compared to another patient in the trial who dies after 2 months of follow-up.

Therefore, standard statistical analyses based on these time-to-first outcome event where subsequent events are ignored may skew the assessment of the treatment effect, lead to biased results, and poorly reflect the true burden of the patient's disease experience (Anker and Mcmurray 2012; Anker et al. 2016; Ferreira-González et al. 2007b; Freemantle et al. 2003; Lubsen and Kirwan 2002; Prieto-Merino et al. 2013; Freemantle et al. 2003; Heddle and Cook 2011; Claggett et al. 2013; Brown et al. 2016). Moreover, the composite endpoint of time-to-first event are not applicable when the component outcomes are on different scales, such as a mixture of discrete, continuous, time-to-event, and quality-of-life outcomes (Felker et al. 2008; Tyler et al. 2011; Bebu and Lachin 2015).

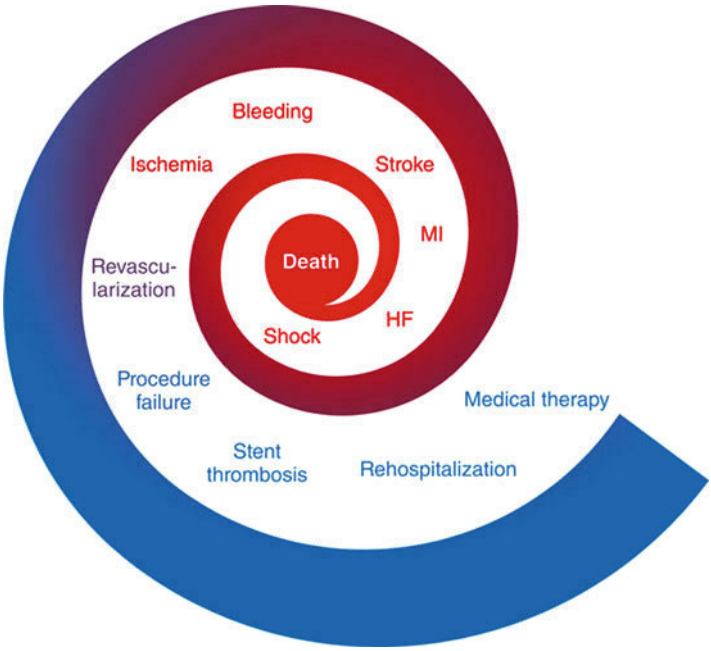


Fig. 1.1 Relative severity of cardiovascular disease outcomes (from death onward out of the spiral). *Acronyms:* MI: myocardial infarction; HF: heart failure (used with permission from Armstrong and Westerhout (2017))

Despite these serious limitations, analyses of composite endpoints are ubiquitous in a large number of clinical research areas including cardiovascular disease (Lisa and James 1997; Bakal et al. 2012b,a, 2015; Neaton et al. 2005; Follmann et al. 1992; Brittain et al. 1997; Felker et al. 2008), infectious diseases (Neaton et al. 1994; Finkelstein and Schoenfeld 1999; Follmann et al. 2007), oncology (Freemantle et al. 2003), nephrology (Hariharan et al. 2003; Li et al. 2001), neurology and psychiatry (Davis et al. 2003), health services, autoimmune disease, dermatology (Kaufman et al. 1998), respiratory (Spencer et al. 2007), rheumatoid arthritis, limb ischemia (Subherwal et al. 2012), orthopedics (DeCoster et al. 1999), urology, anesthesia, migraines, obstetrics, and gynecology (Ross 2007; Wen et al. 2017)—even though their limitations and unsatisfactory characteristics are widely recognized and genuinely mentioned in most publications (Manja et al. 2017; Zhang et al. 1997; Anker et al. 2016; Tyler et al. 2011; Cordoba et al. 2010; Rowan et al. 2008; Prieto-Merino et al. 2013).

Several alternative methods have been proposed to combine multiple outcomes while taking into account their clinical priority (Lisa and James 1997; Bakal et al. 2015; Neaton et al. 2005; Follmann et al. 1992; Brittain et al. 1997; Felker et al. 2008). Among them are the methods based on prioritized outcomes where component outcomes are prioritized and ordered—following a specific, prespecified

hierarchy and with respect to their clinical importance—from the most severe (e.g., mortality) to the least severe one (or more favorable). Usually, the clinical questions of interest dictate the choice and order of the prioritized endpoints. Treatment comparison requires pairwise comparisons of patients' outcomes, where each pair comprise one patient from one treatment group (e.g., active treatment) and another patient from the alternative treatment group (e.g., control treatment). The statistical underpinnings of these methods are based on ranks. These ranks are used to draw inference as to whether a randomly selected patient in the active treatment will have, on average, a better overall composite endpoint compared to a randomly selected patient in the control treatment group by using the Wilcoxon–Mann–Whitney (WMW) test statistic. These methods, which are considered as part of the global rank approaches (Huang et al. 2008; Ramchandani et al. 2016), can be classified into two distinct categories based on the decision rules that dictate how to proceed from one outcome to a subsequent outcome on the hierarchy of outcomes.

On the one hand, we have the *proportion in favor of treatment* (PFT) of Buyse (2010) (also known as the win difference Luo et al. 2017) and the *win ratio* (WR) introduced by Pocock et al. (2011), which follow the ideas from Moyé et al. (1992) and Finkelstein and Schoenfeld (1999). In these methods, pairwise outcome comparisons between patients from the active and control treatment groups are conducted, starting from the most severe outcome. For each pairwise comparison, the patient with a better outcome is declared a winner. If it is not possible to determine the winner (e.g., comparison inconclusive or indeterminate) on the most severe outcome, the two patients are then compared on the second most severe outcome, and so forth. Finally, each patient score is recorded as a win (better outcome in the pairwise comparison), a loss (worse outcome), or a tie (when unable to declare a winner after exhausting all available outcomes).

The PFT is defined as the difference between the proportions of wins in the active and control treatment groups. The null hypothesis of no difference between the treatment groups corresponds to a PFT that is equal to 0, while a positive (resp., negative) value demonstrates that the active treatment is better (worse) than the control treatment. Similarly, the WR is the ratio of the proportion of wins in the active treatment over the proportion of wins in the control treatment. Under the null hypothesis, the WR is equal to 1. It is greater (resp., less) than 1 when the active treatment is beneficial (disadvantageous) compared to the control treatment.

On the other hand, we have the *worst-rank score* analysis—based on the original idea of Gould (1980) and O'Brien (1984). For this method, patients are placed into "buckets" (to use the analogy from Subherwal et al. 2012) on the hierarchy of component outcomes. In other words, each patient is categorized based on her or his worst personally experienced outcome. All the patients who have experienced the worst outcome (e.g., those who died) are assigned to the lowest-ranked bucket, patients who did not experience the worst outcome, but the second worst outcome are placed in the second lowest-ranked bucket, and so forth. Finally, depending on the predetermined choice of the component outcomes, patients with the less severe outcome or who did not experience any of the component outcomes are assigned to the highest-ranked bucket (Lachin 1999; Matsouaka and Betensky 2015; Matsouaka

et al. 2016). Then, every patient in the active treatment group is compared to every patient in the control treatment group to determine whether the actively treated patient's outcome is better than or the same as the outcome of the patient in the control treatment.

The final result is determined by the buckets the compared patients belong to and by their respective outcomes. If the pair of patients is from the same bucket, they are compared by the magnitude of their outcome measures or by their first times to the event (whichever characterizes the bucket), where the longer the time-to-event the better (e.g., later death will be considered better compared to earlier death). If the two patients belong to two different buckets, the patient in the higher-ranked bucket is considered to have a better outcome than the patient in the lower-ranked bucket. Therefore, at the end of the process, all patients are ranked.

Despite the seemingly resemblance between the WR (or the PFT) and the worst-rank score analysis, there are stark clinical and statistical methodological differences between them. Therefore, the choice of one method versus the other must be motivated by the clinical questions of interest and should be predetermined before any analysis. This choice must not be merely dictated by the convenience to pick a method that provides the most significant results. Unlike the win ratio where the focus is put first on the worst outcome and where the next consecutive ranked outcomes (or events experienced by patients) are leveraged only to break ties, with the worst-rank score analysis the first most important step is to place patients in buckets, depending on the worst outcome or event they have personally experienced. Pairwise comparison of patients in one group versus the other is done within and between buckets. When the outcomes of patients from the same bucket are tied, the patients are declared similar and are ranked accordingly. No further comparison is needed. Likewise, when patients are from two different buckets, the patient in the higher-ranked bucket is always considered to have a better outcome.

In practice, the win ratio (or the proportion in favor of the treatment) is used in randomized trials where the most severe outcome is the main outcome of interest. In those trials, it is anticipated that a good percentage of patients will have the most severe outcome, which justify the a priori set to such an outcome. For instance, Pocock et al. (2011) reanalyzed the EMPHASIS-HF data to compare eplerenone against placebo in 2737 patients with NYHA class II heart failure and an ejection fraction less than 35% who were recruited at 278 centers in 29 countries. 1364 patients were randomly assigned to eplerenone and 1373 to placebo and the median follow-up time was 21 months. Pairs of patients from eplerenone and placebo were compared first on cardiovascular (CV) death and, if it was not possible to determine who had a CV death before the other, it was then determined who had a heart failure hospitalization first. Overall, there were a total of 147 deaths (10.8%) in the eplerenone group and 185 (13.5%) in the placebo group attributed to cardiovascular causes. Of the patients receiving eplerenone, 164 (12.0%) were hospitalized for heart failure, as compared with 253 patients (18.4%) receiving placebo.

The worst-rank score analysis is mostly used in trials where the most severe outcome is not the primary outcome. Usually, it is expected that a small percentage of patients will experience the most severe outcome. Therefore, it is mostly used in

settings where the primary interest lies on a nonterminal (nonfatal) outcome, but for which analyses of the observed data are complicated due to the presence of missing observations due to death.

Felker and Maisel (2010) proposed the use of worst-rank score analysis in a hypothetical study of a phase II acute heart failure trial. They suggested a global rank score analysis of 200 patients with 101 patients in the active treatment group and 99 in the placebo group. Patients were compared for in-hospital mortality (4% patients in each group), lack of dyspnea improvement at 24 h (44% patients in active treatment group and 54% in the placebo group), detectable troponin or an increase in troponin by 25% during index hospitalization (7 and 5%, respectively), creatinine increase by more than 0.3 mg/dl (7% and 10%), and finally on change in pro-BNP from randomization to discharge. In another example, Lachin (1999) reexamined a clinical trial of the effect of vesnarinone versus placebo on patients with congestive heart failure and used a worse-rank score analysis of exercise time after 12 weeks of treatment after treatment and death (Feldman et al. 1991). Of the 80 patients randomized (40 in each group), six died before week 12 with five of them in the placebo group.

In this chapter, we consider the worst-rank score analysis and present a framework that allows us to weight the components of a worst-rank (composite) endpoint by relying uniquely on the data at hand. Matsouaka and Betensky studied the statistical properties of the worst-rank analyses based on the (ordinary) Wilcoxon–Mann–Whitney (WMW) test. They considered both tied worst-rank scores (all patients who died are assigned a fixed score) and untied worst-rank scores (where patients who died are ranked based on their time to the death, with the longer time to death the better) in the ranking of the components of the composite outcomes (Matsouaka and Betensky 2015).

For this chapter, we focus on the untied worst-rank score analyses. We assume that we have a data set where we can identify approximately well the time-to-death for each patient who died during the follow-up time. Although, one can easily adapt our method and result in the context of a tied worst-rank analysis. The current framework extends the worst-rank analysis of Matsouaka and Betensky by providing a weighted test statistic where its corresponding weights are optimal in the sense that they maximize the power of the test under a particular alternative hypothesis. We explore the statistical properties of the optimal weighted WMW test on a worst-rank composite endpoint, looking at the null hypothesis of no difference between treatment against a unidirectional alternative hypothesis that the treatment has a favorable effect on the components of the worst-rank composite endpoints or it is at least as effective as the control treatment.

To anchor the framework in the context of worst-rank score analysis, we use, as an example, a randomized clinical trial of acute ischemic stroke conducted at the Massachusetts General Hospital in Boston, Massachusetts. In this trial, a total of 85 patients who had acute ischemic stroke were randomly assigned to either room air (control therapy) or normobaric oxygen therapy (NBO), administered for 8 h. Then, the patients were assessed serially for clinical function scores including the National Institutes of Health stroke scale (NIHSS) score—a function rating scale