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

Lanju Zhang • Ding-Geng (Din) Chen
Hongmei Jiang • Gang Li • Hui Quan
Editors

Contemporary Biostatistics with Biopharmaceutical Applications

Springer

Editors
Lanju Zhang
Data and Statistical Sciences, AbbVie Inc.
North Chicago, IL, USA

Hongmei Jiang
Department of Statistics
Northwestern University
Evanston, IL, USA

Hui Quan
Sanofi US
Bridgewater, NJ, USA

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

Gang Li
R&D
Janssen Pharmaceuticals
Raritan, NJ, USA

ISSN 2199-0980 ISSN 2199-0999 (electronic)
ICSA Book Series in Statistics
ISBN 978-3-030-15309-0 ISBN 978-3-030-15310-6 (eBook)
https://doi.org/10.1007/978-3-030-15310-6

Preface

This book is a collection of the most significant research papers presented at the 26th ICSA Applied Statistics Symposium. Held on June 25–28, 2017, at Hilton Chicago Downtown, this symposium attracted more than 800 statisticians in academia, government, and industry around the world. With the theme *Statistics for a new generation: challenges and opportunities*, the symposium also attracted hundreds of students. One hundred and fifty-nine invited, topic contributed, and contributed sessions covered the broadest variety of topics across the full spectrum of all statistical theoretical fronts and applications. After the symposium, speakers were invited to contribute to this book. From all submissions, the editors selected 18 chapters after rigorous peer-reviewed and subsequent revisions.

The book is organized into two balanced parts: Part I, *Biostatistical Methodology*, which includes nine chapters that present the most recent theoretical breakthrough in experimental design, modelling, and analysis, and Part II, *Biopharmaceutical Applications*, which consists of nine chapters that depict various statistical applications in the biopharmaceutical industry. Each chapter is self-contained with relevant references provided at the end of the chapter. The following is a quick glimpse of each chapter:

Part I. In Chap. 1, Mao developed an EM algorithm to estimate tumor onset time in carcinogenicity studies under the condition that cause of death is unknown in a subset of animals. Log-rank test was used to compare treatment groups against controls. The proposed new method was shown to outperform the available methods by simulation. In Chap. 2, Pan and Jiang addressed the high-dimensional variable selection problem for associating the microbial compositions with a phenotype. They employed a log contrast model to bypass the usual step for normalization and developed a new method to identify phenotype-associated species using penalized regression and stability selection. In Chap. 3, Wei proposed the use of contemporary aggregation as a dimension reduction method in high-dimensional multivariate time series and showed that this natural and simple method had forecast accuracy superior to existing methods. In Chap. 4, Wu developed PC-ABT, a novel principal component-based adaptive-weight burden test for gene-based association mapping

of quantitative traits. This method efficiently accounted for correlation in multiple genotypic variants and was showed to be more powerful than other multiple variant tests that allowed related individuals. In Chap. 5, Wang et al. proposed an adaptive dynamic Bayesian network model that provided an unprecedented tool to elucidate a comprehensive picture of gene regulatory networks. In particular, an unevenly spaced gene expression record can be accommodated. By analyzing real data sets from a surgical study and through extensive simulation studies, the new model was demonstrated for its usefulness and utility. In Chap. 6, Alghamdi et al. focused on ultrafast functional brain imaging studies and proposed an efficient approach for obtaining high-quality stimulus sequences by taking the uncertainty of the autocorrelation of the response into account. The performance of their proposed approach was demonstrated via case studies. In Chap. 7, Zhang et al. derived a global optimization algorithm that provided a guaranteed ε-global optimum for a sparse mixed membership matrix factorization problem. The algorithm was tested on simulated data and small real gene expression data set and found to always bound the global optimum across random initializations and explore multiple modes efficiently. In Chap. 8, Chuang and Yang proposed a nonnegative robust linear model (NRLM) approach that yielded robust, yet interpretable, mixing rate estimates. In a simulation study, NRLM showed a robust performance for finding the relative abundance of specified components when a large amount of noise was present. More importantly, the approach accurately estimated the absolute level of the specified components in the presence of unspecified ones. Finally, it showed a superior performance when applied to deep deconvolution of blood samples. In Chap. 9, O'Brien and Silcox explored optimal experimental designs for parallelism testing in potency bioassays. They derived theoretical optimal designs and proposed several extensions that took practicality into account. One of the designs, reflection design, was demonstrated to be the most efficient and easy to implement since the researcher could merely sketch the drug/compound logistic curves and read off design at some cutoff lines.

Part II. In Chap. 10, Zhang et al. proposed an optimized two-stage phase III clinical trial design that combined three adaptive techniques to offer the opportunity of dose selection and sample size determination based on the first-stage data with strict type I error rate control and robust power across an effect size interval. In Chap. 11, Gou and Chen proposed a generalized framework for critical boundary refinement when conducting hierarchical hypothesis test in a clinical trial involving multiple interim stages with an improvement on the secondary boundary. The framework had a particular advantage when the primary endpoint data can be assessed earlier than the secondary endpoint data. The framework was also extended to include an adaptive update on the refined boundary when the attained sample sizes were different from what they were originally planned. In Chap. 12, Liu et al. proposed an escalation with overdose control design for phase I oncology trials using dose-limiting toxicity (DLT) with two components, one for immediate toxicity in a binary model and the other for late-onset toxicity in a time to event model. They demonstrated that the proposed dose escalation design can incorporate historical

knowledge, protect patients from being assigned to toxic doses, and consider early- and late-onset toxicity while maintaining the escalation timeline. In Chap. 13, Yang et al. proposed a new approach via adding a companion constancy test to the non-inferiority test that consequently protected the validity of a non-inferiority trial under Bayesian framework. In addition, historical data of the active control was borrowed in the analysis with two different approaches. In Chap. 14, Lin et al. introduced a nonparametric model which was robust to event time distribution in response-adaptive designs for survival trials. The operating characteristics of the proposed design and the parametric design were compared by simulation studies, including their robustness properties, with respect to model misspecifications. Both advantages and disadvantages of adaptive randomization were discussed in the summary from a practical perspective of clinical trials as well as illustrations by master protocol case studies. In Chap. 15, Lu et al. provided valuable considerations of the design and analysis of the non-randomized studies using the propensity score methodology. Statistical and regulatory perspectives were highlighted. In Chap. 16, Jiang et al. reviewed key methodological and statistical implications of pragmatic clinical trials (PCTs) in the context of drug development and reimbursement, with emphasis on study design and analyses to maximize external validity. The principles of PCTs challenged some well-established guidelines in randomized clinical trials (RCTs), as open-label and treatment switch in intention-to-treat (ITT) population being the most pronounced ones. They provided valuable suggestions on handling these issues. In Chap. 17, Lipkovich et al. enhanced existing SIDES and SIDEScreen methods for biomarker discovery by incorporating stochastic elements in computing the variable importance, expected treatment effect, and replicability index. The improvement was particularly useful when dealing with relatively small data sets, so as to properly account for the uncertainty of the subgroup selection process. The operating characteristics of the Stochastic SIDEScreen were demonstrated to be improved compared with the corresponding deterministic procedure through simulation. Last, but not the least, in Chap. 18, Pantoja-Galicia and Gene Pennello discussed the implicit or explicit trade-offs between false-positive and false-negative test errors provided by the information from the receiver operating characteristic (ROC) curve. They demonstrated how it can impact the evaluation of the performance of a new medical diagnostic test in comparison with an already established test. They illustrated the idea with the comparability of a new test N with respect to a standard test S in terms of the seriousness of a false-positive error relative to a false-negative error using the information from the ROC curve.

The editors are grateful to many people who contributed to the publication of this book. First, we would like to thank the authors of all chapters for their original research and dedication to share through this book. Second, our sincere appreciations go to all the reviewers for their valuable time and excellent review, which significantly improved the presentations and quality of the book. Third, our gratitude goes to the leadership of the executive committee, organization committees, and numerous volunteers of the 26th ICSA Applied Statistics Symposium. This book would not be possible without such a successful symposium. Last, but not least,

we would like to acknowledge the support and guidance of Abitha PradeepCoumar, Susan Westendorf, and Shobha Karuppiah from Springer through the entire process of publishing this book.

North Chicago, IL, USA — Lanju Zhang
Chapel Hill, NC, USA — Ding-Geng (Din) Chen
Evanston, IL, USA — Hongmei Jiang
Raritan, NJ, USA — Gang Li
Bridgewater, NJ, USA — Hui Quan

Contents