Preface

Researchers often use observational data to estimate treatment effects because randomized controlled trials or experimental designs are not always feasible. The use of observational data poses a threat to the validity of causal inference by introducing selection bias into the treatment assignment. To tackle this problem, in 1983, Paul R. Rosenbaum and Donald B. Rubin theorized propensity score analysis, which balances the distributions of observed covariates between the treatment and control groups, as a means of mimicking characteristics of randomized controlled trials and, therefore, reducing selection bias. Over the past three decades, propensity score methods have become increasingly popular in the social, behavioral, and health sciences for making causal inferences based on observational studies.

However, both methodological and practical challenges persist in the use of propensity score analysis. These include what estimation method should be used to estimate propensity scores, under what conditions propensity score matching is appropriate, how to assess matching quality, how to implement propensity score analysis effectively on complex data, what should be considered after implementing propensity score analysis, and what computer program is appropriate for a specific data condition. This edited book introduces new methodological developments that address these challenges encountered in practice. We do not intend to solve all the challenges in this single volume but rather hope to stimulate

further discussions and, therefore, advance its use. The book is written for statisticians, methodological and practical researchers, and graduate students preparing for careers as statisticians and researchers in the social, behavioral, and health sciences.

The volume includes 15 chapters organized in five parts. Beginning with fundamentals of propensity score analysis, Part I includes an overview of propensity score analysis underlying concepts and current issues (Chapter 1), as well as a review of computer programs for propensity score analysis (Chapter 2). The next three parts encompass the main body of the book, introducing new developments that address current issues in propensity score estimation and matching in Part II (Chapters 3–5), outcome analysis after matching in Part III (Chapters 6–8), and use of propensity score analysis on complex data in Part IV (Chapters 9–11). The last section, Part V, includes discussions on missing data in propensity scores (Chapter 12), sensitivity of unobserved confounding in propensity score analysis (Chapters 13 and 14), and the extension of propensity score analysis to prognostic score analysis for causal inference (Chapter 15).

Chapter 1 provides a comprehensive overview of propensity score analysis and the basic concepts of propensity score methods. By summarizing the existing propensity score matching methods, the steps in propensity score analysis, analysis after matching, and the current issues, Wei Pan and Haiyan Bai present a complete set of road maps for researchers to understand the fundamentals of propensity score analysis. They also build connections for readers between these concepts and the other chapters for further exploration of specific topics of interest.

Modern technology makes statistical computing a less demanding task. Positioned for both methodological and practical researchers, this book allows an overall understanding of the availability of statistical programs that can be used for propensity score analysis. In Chapter 2, Megan Schuler reviews the range of statistical software from which researchers will be able to find one that is appropriate and familiar to conduct propensity score analysis.

Providing that the strong ignorability assumption stands, propensity score as a single score is used to balance the selection bias between the treatment and control groups. It is essential to understand the procedures for propensity score estimation and its new developments to solve some critical problems in propensity score estimation. In Chapter 3, Lane F. Burgette, Daniel F. McCaffrey, and Beth Ann Griffin describe the propensity score estimation steps and provide an illustrative example of how to estimate propensity scores using generalized boosted models.

Propensity score matching is one of the most important propensity score methods; however, it is usually a daunting task for practical researchers to identify which matching method works best for their data. Researchers often struggle with deciding if matching should be done with replacement, how much common support is sufficient, and whether sample size ratio is influential for bias reduction results. Haiyan Bai, in Chapter 4, helps address these methodological considerations in implementing propensity score matching by providing a systematic investigation with 60 experimental trials to assist researchers in selecting an appropriate matching method, and to aid methodologists in initiating further investigations into these intricate issues.

Evaluating the quality of covariate balance is a critical procedure for verifying the validity of using a propensity score method. In Chapter 5, Cassandra W. Pattanayak focuses on how to evaluate balance results on key background covariates after matching. She demonstrates best practices for evaluating balance and addresses the related issues of missing data, generalizability, and an occasional need for using matching or subclassification in randomized experiments.

In Chapter 6, M. H. Clark discusses how to use propensity score methods (or propensity score adjustment methods) to reduce selection bias in quasi-experimental studies. Through both a literature review and a case study, she describes and compares four propensity score adjustment methods in their capability to balance covariates and reduce selection bias in treatment effects for both continuous and categorical outcome variables. She provides guidance and useful information for researchers to understand the conditions under which each method is most appropriate.

Chapter 7 describes an alternative propensity score method, the matching weight. This method was developed to reduce the complexity of using the existing propensity score matching methods. Liang Li, Tom H. Greene, and Brian C. Sauer illustrate how the matching weight method is simple to implement and generally more efficient than the pair matching estimator, while producing accurate variance estimation, improved covariate balance, and double robust estimation.

Chapter 8 focuses on the design and analysis of matched pairs in observational studies. Scott F. Kosten, Joseph W. McKean, and Bradley E. Huitema provide a general overview of the role of propensity scores in the design and analysis of observational studies. They also introduce a newly developed method of outcome analysis and provide an example to illustrate the advantages of the new method over competing approaches. Input

and output for new software that computes this analysis, and competing alternatives are also provided.

In this era of easier access to data, researchers are increasingly faced with more complex datasets that are challenging to analyze. In Chapter 9, Walter L. Leite offers an overview of the use of propensity score matching methods in longitudinal studies. He presents a set of growth models that can be used to analyze the matched longitudinal dataset. He demonstrates the application of the methods for researchers to better understand how to apply propensity score methods to longitudinal data and the use of statistical programs in such situations. In Chapter 10, Qiu Wang provides researchers with useful and critical information for applying propensity score matching methods to clustered or multilevel data. He introduces the dual-matching method for reducing selection bias due to both individual and cluster covariates when clusters are the intervention units.

Researchers may be engaged in analyzing data from complex samples, rather than a simple random sample, such as with national or international surveys. In Chapter 11, Debbie L. Hahs-Vaughn provides a general review of the literature on this topic and enables readers to understand the applications of propensity score analysis to complex samples and the impact of ignoring the sampling design on parameter estimation.

Issues of missing data cannot be ignored in data analysis. In propensity score analysis, missing data can significantly influence propensity score estimation. In Chapter 12, Robin Mitra addresses the problem of missing covariate data and the existing imputation approaches. Multiple imputation procedures are illustrated for researchers to understand possible ways to handle missing covariates in propensity score estimation.

Evaluation of the impact of potential uncontrolled confounding is an important component for causal inference in observational studies: the propensity score methods do not deal with unobserved confounders. In Chapter 13, Rolf H. H. Groenwold and Olaf H. Klungel introduce propensity score calibration for directly controlling for unobserved confounding and discuss the extent to which propensity score methods indirectly control for unobserved confounding. They also illustrate how the sensitivity analysis can be used as a diagnostic tool for unobserved confounding. In Chapter 14, Lingling Li, Changyu Shen, and Xiaochun Li introduce two additional sensitivity analysis approaches that are based on the inverse probability weighting estimator, but extend it to accounting for uncontrolled confounding from an alternative perspective. In addition, they demonstrate a real-life example for applying the two sensitivity analyses.

It is always helpful for researchers to know extensions to the existing methods. In Chapter 15, Ben Kelcey and Christopher M. Swoboda introduce prognostic score analysis as a complement to propensity score analysis and applications of prognostic score analysis to clustered or multilevel settings. They also illustrate the use of prognostic score analysis using an empirical example to investigate the treatment effects. This last chapter exposes readers to an extension of propensity score analysis in an effort to keep a wide vision toward new approaches.

The chapters are logically ordered with basic contents presented in the early chapters. Thus, the fundamentals part (Part I) assists readers with limited knowledge of or experience with propensity score analysis to learn basic concepts and understand the current issues in propensity score analysis. Readers who have mastered the fundamentals and those with some prior knowledge of propensity score analysis can benefit from reading the later chapters for cutting-edge discussions about developments and extensions in propensity score analysis. The chapters also can be used independently because each one focuses on a specific topic. This book will serve as an excellent supplemental text for advanced research methods courses.

This book has both methodological and practical value. Methodologists can use this book to further explore new directions and improvement of propensity score methods. Also, it can help practical researchers to use propensity score analysis both appropriately and directly to improve the validity of their research on causal claims. In addition, this book provides necessary statistical program codes and application examples for practical researchers to follow.

Given that this is our first edited book, we greatly appreciated the chapter authors' extensive involvement. Without their remarkable cooperation, flexibility, and dedication, this book would not be possible. They not only contributed their own excellent chapters to this book, but many also spent their precious time peer-reviewing others' chapters. Along with three ad hoc chapter reviewers who kindly lent their expertise to the project, the chapter reviewers were:

Lane F. Burgette, RAND Corporation

Haiwen Chen, Global Assessment Department, Educational Testing Service

M. H. Clark, Department of Educational and Human Sciences, University of Central Florida

Beth Ann Griffin, RAND Corporation

Rolf H. H. Groenwold, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

Debbie L. Hahs-Vaughn, Department of Educational and Human Sciences, University of Central Florida

Walter L. Leite, Research and Evaluation Methodology Program, University of Florida

Liang Li, Department of Biostatistics, The University of Texas MD Anderson Cancer Center

Joseph W. McKean, Department of Statistics, Western Michigan University

Daniel F. McCaffrey, Educational Testing Service

Robin Mitra, Mathematical Sciences, University of Southampton, United Kingdom

Megan Schuler, Department of Mental Health, Johns Hopkins Bloomberg School of Public Health

Elizabeth A. Stuart, Departments of Mental Health and Biostatistics, Johns Hopkins Bloomberg School of Public Health

Christopher M. Swoboda, School of Education, University of Cincinnati

Qiu Wang, Department of Higher Education, Syracuse University

Rebecca Zwick, Statistical and Psychometric Theory and Practice, Educational Testing Service