

# Preface

This book originated from a series of discussions among the editors when we were all at the University of Rochester, NY, before 2015. At that time, we had a research discussion group under the leadership of Professor Xin M. Tu that met biweekly to discuss the methodological development on statistical causal inferences and their applications to public health data. In this group, we got a closer overview of the principles and methods behind the statistical causal inferences which are needed to be disseminated to aid the further development in the area of public health research. We were convinced that this can be accomplished better through the compilation of a book in this area.

This book compiles and presents new developments in statistical causal inference. Data and computer programs will be publicly available in order for readers to replicate model development and data analysis presented in each chapter so that these new methods can be readily applied by interested readers in their research.

The book strives to bring together experts engaged in causal inference research to present and discuss recent issues in causal inference methodological development as well as applications. The book is timely and has high potential to impact model development and data analyses of causal inference across a wide spectrum of analysts, as well as fostering more research in this direction.

The book consists of four parts which are presented in 15 chapters. Part I includes Chap. 1 with an overview on statistical causal inferences. This chapter introduces the concept of potential outcomes and its application to causal inference as well as the basic concepts, models, and assumptions in causal inference.

Part II discusses propensity score method for causal inference which includes six chapters from Chaps. 2 to 7. Chapter 2 gives an overview of propensity score methods with underlying assumptions for using propensity score, and Chap. 3 addresses causal inference within Dawid's decision-theoretic framework, where studies of "sufficient covariates" and their properties are essential. In addition, this chapter investigates the augmented inverse probability weighted (AIPW) estimator, which is a combination of a response model and a propensity model. It is found that, in the linear regression with homoscedasticity, propensity variable analysis provides exactly the same estimated causal effect as that from multivariate linear regression,

for both population and sample. The AIPW estimator has the property of “double robustness,” and it is possible to improve the precision given that the propensity model is correctly specified.

As a critical component of propensity score analysis to reduce selection bias, propensity score estimation can only account for observed covariates, and this estimation to unobserved covariates has not been fully understood. Chapter 4 is then designed to introduce a new technique to assess the robustness of propensity score estimation methods to unobserved covariates. A real dataset on substance abuse prevention for high-risk youth is used to illustrate this technique.

Chapter 5 discusses the missing confounder data in propensity score methods for causal inference. It is well known that the propensity score methods, including weighting, matching, or stratification, have been used to control potential confounding effects in observational studies and non-randomized trials to obtain causal effects of treatment or intervention. However, there are few studies to investigate the missing confounder data problem in propensity score estimation which is unique and different from most missing covariate data problem where the goal is parameter estimation. This chapter is then to review and compare existing methods to deal with missing confounder data in propensity score methods and suggest diagnostic checking tools to select a suitable method in practice. In Chap. 6, the focus is turned to the models of propensity scores for different kinds of treatment variables. This chapter gives a thorough discussion of all methods with a comparison between parametric and nonparametric approaches illustrated by a public health dataset. Chapter 7 is to discuss the computational barrier in propensity score in the era of big data with example in optimal pair matching and consequently offer a novel solution by constructing a stratification tree based on exact matching and propensity scores.

Part III is designed for causal inference in randomized clinical studies which includes five chapters from Chaps. 8 to 12. Chapter 8 reviews important aspects of semiparametric theory and empirical processes that arise in causal inference problems with discussions on empirical process theory, which provides powerful tools for understanding the asymptotic behavior of semiparametric estimators that depend on flexible nonparametric estimators of nuisance functions. This chapter concludes by examining related extensions and future directions for work in semiparametric causal inference.

Chapter 9 discusses the structural nested models for cluster-randomized trials for clinical trials and epidemiologic studies. It is known that in clinical trials and epidemiologic studies, adherence to the assigned components is not always perfect. In this chapter, the estimation of causal effect of cluster-level adherence on an individual-level outcome is provided with two different methodologies based on ordinary and weighted structural nested models (SNMs) which are validated by simulation studies. The methods are then applied to a school-based water, sanitation, and hygiene study to estimate the causal effect of increased adherence to intervention components on student absenteeism. In Chap. 10, the causal models for randomized trials with two active treatments and continuous compliance are addressed by first proposing a structural model for the principal effects and

then specifying compliance models within each arm of the study. The proposed methodology is illustrated with an analysis of data from a smoking cessation trial.

In Chap. 11, the causal ensembles for evaluating the effect of delayed switch to second-line antiretroviral regimens are proposed to deal with the challenge in randomized clinical trials of delayed switch. The method is applied for cohort studies where decisions to switch to subsequent antiretroviral regimens were left to study participants and their providers as seen from ACTG 5095. Chapter 12 is to introduce a new class of structural functional response models (SFRMs) in causal inference, especially focusing on estimating causal treatment effect in complex intervention design. SFRM is an extended version of existing structural mean models (SMMs) that is widely used in the area of randomized controlled trials to provide optimal solution in estimation of exposure-effect relationship when treatment exposure is imperfect and inconsistent to every individual subject. With a flexible model structure, SFRM is ready to address the limitations of existing approaches in causal inference when the study design contains multiple intervention layers or dynamic intervention layers and capable to offer robust inference with a simple and straightforward algorithm.

Part IV is devoted to the structural equation modeling for mediation analysis which includes three chapters from Chaps. 13 to 15. In Chap. 13, the identification of causal mediation models with an unobserved pretreatment confounder is explored on identifiability of mediation, direct, and indirect effects of treatment on outcome. The mediation effects are represented by a causal mediation model which includes an unobserved confounder, and the direct and indirect effects are represented by the mediation effects. Simulation studies demonstrate satisfactory estimation performance compared to the standard mediation approach. In Chap. 14, the causal mediation analysis with multilevel data and interference is studied since this type of data is a challenge for causal inference using the potential outcomes framework because the number of potential outcomes becomes unmanageable. Then the goal of this chapter is to extend recent developments in causal inference research with multilevel data and violations of the interference assumption to the context of mediation. This book concludes with Chap. 15 to compressively examine the causal mediation analysis using structure equation modeling by taking advantage of its flexibility as a powerful technique for causal mediation analysis.

As a general note, the references for each chapter are at the end of the chapter so that the readers can readily refer to the chapter under discussion. Thus each chapter is self-contained.

We would like to express our gratitude to many individuals. First, thanks go to Professors Xin M. Tu and Wan Tang for leading and organizing the research discussion which led the production of this book. Thanks go to Hannah Bracken, the associate editor in statistics from Springer; to Jeffrey Taub, project coordinator from Springer (<http://link.springer.com>); and to Professor Jiahua Chen, the coeditor of Springer/ICSA Book Series in Statistics (<http://www.springer.com/series/13402>), for their professional support of the book. Special thanks are due to the authors of the chapters.

We welcome any comments and suggestions on typos, errors, and future improvements about this book. Please contact Professor Hua He (hhe2@tulane.edu), Pan Wu (PWu@ChristianaCare.org), or Ding-Geng (Din) Chen (DrDG.Chen@gmail.com or dinchen@email.unc.edu).

New Orleans, LA, USA  
Newark, DE, USA  
Chapel Hill, NC, USA  
March 2016

Hua He, Ph.D.  
Pan Wu, Ph.D.  
Ding-Geng (Din) Chen, Ph.D.

Statistical Causal Inferences and Their Applications in  
Public Health Research

He, H.; Wu, P.; Chen, D.D.-G. (Eds.)

2016, XV, 321 p. 24 illus., 11 illus. in color., Hardcover

ISBN: 978-3-319-41257-3