

Editorial: Special Issue on “Causal Inference”

Jason Roy and Dylan Small

1. INTRODUCTION

Understanding cause and effect is central to many scientific investigations and public policy questions, and scientific papers often use statistical analyses in arguing for or against cause and effect relationships. Early important statisticians such as R. A. Fisher played an important role in the development of randomized trials which are one of the most powerful tools for understanding cause and effect. Yet outside of ideal randomized trials, statisticians have traditionally been wary of engaging in discussions of cause and effect. In his 1986 paper “Statistics and Causal Inference,” Paul Holland opened with, “The reaction of many statisticians when confronted with the possibility that their profession might contribute to a discussion of causation is immediately to deny that there is any such possibility.” Times have changed. In recent years, there has been an explosion of interest in statistics in causal inference.

Major advances over the past few decades have changed the field drastically. These include definitions of causal effects for complex questions involving dynamic treatment regimes, mediation, interference, or local average treatment effects; their corresponding identifying assumptions; causal graphs; methods for carrying out sensitivity analysis; new study designs; and developments in semi-parametric theory and machine learning, to name a few. These advances have gone from initial publication in the statistical and related literature to widespread application impacting just about every field. There are now conferences, courses and journals devoted to causal inference research.

The field of causal inference is still evolving. Technological advances are opening up new possibilities, with more data being collected on just about everything, along with faster computing. There are opportunities to pursue questions that were not previously possible, and to answer questions in new ways. This special issue highlights many of these, including topics involving big data, machine learning, micro-randomized trials. However, despite

the advances in technology, the field of causal inference is never fully black-box. Proper causal inference necessarily requires careful thought about the problem, as identification relies on uncheckable assumption. This special issue also includes important work that highlights limitations of what we can know.

The issue includes papers from leading experts in causal inference who have a variety of perspectives on the topic. Each article also includes discussants and rejoinder.

2. SUMMARY OF ARTICLES

Matching is a popular approach to address, and correct for, nonrandom treatment assignment. It is well known that optimal matching is computationally intensive, making it impractical to implement for large datasets. “Matching Methods for Observational Studies Derived from Large Administrative Databases” by Yu, Silber and Rosenbaum proposes a new approach to address this issue. Their approach involves using an iterative algorithm to find an optimal propensity score caliper, which reduces the number of candidate matches. They propose further refinements that can make optimal matching feasible, including restricting the number of near neighbors or imposing near-fine balance constraining. The methodology is illustrated using a Medicaid data set involving more than 150,000 potential controls. The discussants raise many interesting points, including about the use of calipers, comparisons between full and pair matching, and challenges with EHR data.

Traditionally, randomized trials have assigned people to one of two or more possible treatments at baseline and then followed them up. Smartphones offer the possibility of rapidly adapting a person’s treatment based on a person’s dynamic responses, but the traditional randomized trial cannot be used to compare dynamic treatment strategies. Micro-randomized trials are a recently developed type of randomized trial in which the treatments are randomized numerous times for each individual. As in many causal inference settings, we would like to not only understand average treatment effects but treatment effect heterogeneity so as to better be able to personalize treatments. The paper “Linear mixed models under endogeneity: modeling sequential treatment effects with application to a mobile health study” by Qian, Klasnja and Murphy consider how to make inferences about treatment effect heterogeneity in a micro-randomized trial.

Jason Roy is Professor of Biostatistics, Rutgers School of Public Health, Piscataway, New Jersey 08854, USA (e-mail: jason.roy@rutgers.edu). Dylan Small is Class of 1965 Wharton Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: dsmall@wharton.upenn.edu).

Linear mixed models are one way to model treatment effect heterogeneity but directly applying them in a micro-randomized trial could provide biased estimates because potential moderators of the treatment may depend on prior treatment and be endogenous. Qian et al. provide a careful discussion of the conditions under which such bias would arise and show that under a plausible conditional independence assumption, valid estimates of causal effects that are conditional-on-the-random effect can be obtained. Qian et al.'s results illuminate when standard linear mixed model software can be used effectively in analyzing treatment effect heterogeneity in micro-randomized trials and when it will instead be biased. The three discussants offer interesting perspectives on many points, including understanding and testing the conditional independence assumption that Qian et al. consider.

In the paper "Invariance, Causality and Robustness," which is an expanded version of the 2018 Neyman Lecture, Buhlmann considers causal inference from a different perspective from most of the other papers in this issue. The paper summarizes much of the recent work on invariance of associations across environments and viewing causal effects as optimizing worst case risk. Important new work on linear and nonlinear anchor regression (related to instrumental variable methods) is developed. Both of the discussants offer interesting perspectives on the potential outcomes approach versus invariance-based approach for defining and identifying causal effects.

Most empirical causal studies in the scientific literature focus on one or a narrow range of outcomes. For example, a paper might examine the effect of being physically abused as a child on the risk of having a heart attack as an adult or the risk of several cardiovascular outcomes such as heart attack, stroke etc. Another paper might examine the effect of being physically abused as a child on the risk of developing Alzheimer's disease and another on a person's earnings as an adult. In their paper "Outcome-wide longitudinal designs for causal inference: a new template for empirical studies," VanderWeele, Mathur and Chen (2020) propose that when studying the effect of an exposure, instead of just focusing on one or a few related outcomes at a time, researchers should consider a wide range of outcomes. They discuss advantages of doing so, which include conveying information about the effects of an exposure accurately and concisely, efficiently using research resources and lessening publication bias. VanderWeele et al. develop a comprehensive new design for studying the causal effects of an exposure on a wide range of outcomes, including methods for confounder selection, multiple testing and sensitivity analysis. The discussants present interesting perspectives on robust methods of confounder selection, accounting for missing data and individualized treatment strategies when considering VanderWeele et al.'s suggested design of looking at the effects of an exposure on a wide range of outcomes.

Randomized trials enable causal inference because the probability of treatment assignment is known and thus the null distribution under a causal hypothesis can be found (Fisher, 1935). Many estimators for causal inference for observational studies are based on assuming that all confounders are measured, estimating the probability of treatment assignment given the confounders (the propensity score) and then treating the study like a randomized trial in which the treatment assignment is known (Rosenbaum and Rubin, 1983). A key assumption needed for such estimators to consistently estimate the average treatment effect is positivity (also called overlap), which means everyone in the population of interest has a probability of treatment that is strictly between 0 and 1. Even when positivity holds, near-positivity violations, that is, some people have a probability of treatment given their confounders that is close to 0 or 1, are often present and can cause estimators to be erratic. In their paper "A nonparametric super-efficient estimator of the average treatment effect," Benkeser, Cai and Van der Laan develop two new estimation techniques that perform less erratically than existing estimators and perform well whether or not near positivity violations are present. The discussants provide interesting perspectives on the proper role of automation in causal inference, stabilizing propensity scores and what is needed to make the new estimators practically usable among other topics.

The final article in this special issue covers fundamental issues regarding accuracy of confidence intervals in some causal inference settings. A lot of recent work on causal inference has included doubly robust estimators that utilize prediction models developed from the machine learning literature. For Wald confidence intervals of causal effects from such approaches to be accurate, the bias of the estimator should be of smaller order than the standard error. In their paper "On nearly assumption-free tests of nominal confidence interval coverage for causal parameters estimated by machine learning," Liu, Mukherjee and Robins propose a way to test whether that is indeed the case for parameters in a particular class. In addition to the theoretical work, they also illustrate some of the key ideas using simulation studies. We expect that this paper will also lay the foundation for future work addressing fundamental questions about what we can learn about causal parameters with minimal assumptions. The discussants expand on the ideas, exploring questions around covariate structure, parametric inference, and structure- versus methods-driven inference.

We hope that readers will enjoy the articles and discussions, and get a sense for the breadth of issues that are actively being explored in the field of causal inference. We also hope it will stimulate new research ideas and collaborations, leading to many exciting discoveries and advances over the next decade.

REFERENCES

- FISHER, R. A. (1935). *The Design of Experiments*. Hafner, New York.
- HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–970. MR0867618
- ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 <https://doi.org/10.1093/biomet/70.1.41>