Statistical Science
2020, Vol. 35, No. 3, 335–337
https://doi.org/10.1214/20-STS800

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 Casual 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 semiparametric 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).