



Award Details

Improving Causal Inference Methods in Statistics for Analyzing Big Data

Research Details

Competition Year:	2018	Fiscal Year:	2018-2019
Project Lead Name:	Karim, Mohammad	Institution:	University of British Columbia
Department:	Medicine, Faculty of	Province:	British Columbia
Award Amount:	21,000	Installment:	1 - 5
Program:	Discovery Grants Program - Individual	Selection Committee:	Mathematics and Statistics
Research Subject:	Biostatistics	Area of Application:	Medical and health sciences
Co-Researchers:	No Co-Researcher	Partners:	No Partners

Award Summary

The increasing availability, declining cost of computational machineries and wider application of smart and cloud-based technologies have led to a growing trend of collecting large-scale information for business, utilitarian and scientific purposes. These databases generally contain a considerable number of variables, cover substantially large populations with long follow-up, and better reflect 'real-world' daily practices compared to those derived from carefully controlled randomized experiments. However, these datasets are not primarily collected for research purposes, and in the absence of randomization, confounding poses a critical challenge in exploring the cause-and-effect relationship between the outcome and the intervention. There is a vast literature on confounding adjustment in the statistical and causal inference literature that guides us to select appropriate variables to adjust and control, e.g., controlling for confounders and risk factors, but not adjusting for instruments and noise variables. Due to the complexity and large size of these databases with thousands of variables, it is not tenable for a domain expert to (i) hand-pick the important confounders or identify which variables are instruments, (ii) reasonably correctly guess the functional form of the covariates in the intervention model (in the propensity score context) or the outcome model, (iii) adequately assess the covariate balance for so many variables. *****To address these challenges, there are four specific research objectives in this proposal. 1. To develop confounder selection approaches in a high dimensional setting incorporating the principles established in the causal inference literature. 2. To study the robustness of various data-adaptive methods in the context of model misspecification in a high dimensional setting. 3. To propose appropriate metrics for assessing the 'covariate balance' in the context of propensity scores estimated from high-dimensional covariates. 4. To investigate the above issues when longitudinal data are available. These methods will be evaluated through theoretical developments, real-life applications, and via realistic simulations. *****I am positioned in a unique interdisciplinary research environment, as an Assistant Professor in the UBC Faculty of Medicine, a biostatistician at St. Paul's hospital, and an alumnus from the Statistics department, UBC, with close research ties with UBC and McGill. In this big-data era, there are huge demands for students with training in statistical modeling who can take causal structures into consideration while analyzing a large data set. Training of highly qualified personnel within an interdisciplinary environment is an essential component of this research. Trainees will receive training and access to high-quality research datasets and methodological and applied research questions that will have a real-life impact.