JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Department of Biostatistics

# BIOSTATISTICS SEMINAR,

## A Machine Learning Approach to Causal Inference in the Presence of Missing Data

Xiaochun Li
Department of Biostatistics, School of Medicine, Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana 46202, U.S.A.

Abstract:

Observational medical databases increasingly find uses for comparative effectiveness and safety research. However, the lack of analytic methods that simultaneously handle the issues of missing data and confounding bias along with the onus of model specification, limit the use of these valuable data sources. We derived a novel machine-learning approach based on trees to estimate the average treatment effect. In order to evaluate causal estimation by model-free machine-learning methods in data with incomplete observations, we conducted a simulation study with data generated from known models of exposure, outcome and missing mechanisms. Thus, the true causal effect was known and used as the benchmark for evaluations. Two settings were studied. We compared the bias and standard error of causal estimates from our method to a multiply robust parametric method, the complete case analysis (CC) and a regression analysis after multiple imputations (MI). The proposed methods were applied to a real observational data set of implantable cardioverter defibrillator use.

Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics
Monday, February 18, 2019, 12:15:1-15, Room W2008 (Refreshments 12:00pm)

For disability access information or listening devices, please contact the Office of Support Services at 410-955-1197 or on the web **www.jhsph.edu/SupportServices, EO/AA**