

THE GRADUATE COLLEGE OF THE
UNIVERSITY OF OKLAHOMA HEALTH SCIENCES CENTER

ANNOUNCES THE FINAL EXAMINATION OF

Michael Machiorlatti

FOR THE DEFENSE OF THE DOCTOR OF PHILOSOPHY DEGREE GRADUATE COLLEGE

Department of Biostatistics and Epidemiology

Friday, November 22, 2019 | 12:30 pm

College of Public Health, Room 144



EXPLORATION OF SELECTION BIAS, THE HECKMAN CORRECTION AND MISSING DATA IMPUTATION

COMMITTEE IN CHARGE: Dr. Sixia Chen, Ph.D., Co-Chair, Dr. Sara Vesely, Ph.D., Co-Chair, Dr. Janis Campbell, PhD, Dr. Kai Ding, PhD, and Dr. James George, MD.

ABSTRACT: The Heckman selection model proposed by Heckman (1979) has been used extensively in economics and social sciences to correct selection bias. It is closely related to nonignorable nonresponse problems. Inference based on the Heckman selection model with complex survey data has not been rigorously studied. Results from the Heckman model, which are attained by assuming simple random sampling, can be biased and misleading when used in a complex survey. We studied properties of estimates from Heckman model by incorporating sampling design features. In addition, we propose efficient weight smoothing approaches to further improve efficiency of the estimates under informative sampling. Simulation studies show the benefits of our proposed methods. Our proposed estimators have smaller bias and mean square error than the design-based estimator. An empirical study using (2015-16) NHANES was then explored utilizing our techniques in order to understand their implications using a real data application.

Missing data is an issue across all fields that involve data analysis. Methods for missing data analysis has been explored in broad settings in order to understand how the missing data influences the efficiency and bias of certain estimators and techniques. In this research, we explore parametric and non-parametric imputation procedures in order to understand their effectiveness. We then employ a doubly robust algorithm comprised of model and propensity score information using a purely nonparametric approach utilizing regression trees to explore the benefits of this method over utilizing one singular imputation approach. Specifically, we explore how regression imputation, multiple imputation, predictive mean matching, hot deck imputation, nearest neighbor and classification trees perform in a simulated missing scenario. Our study departs from traditional imputation papers in that we simulate missingness from NHANES 2015-16 survey in order to evaluate the procedures in a real data scenario with complex survey design.