
STATISTICS SEMINAR

UW-Department of Statistics

www.stat.wisc.edu



Abstract: Randomized controlled trials (RCTs) are increasingly prevalent in education research, and are often regarded as a gold standard of causal inference. Two main virtues of randomized experiments are that they (1) do not suffer from confounding, thereby allowing for an unbiased estimate of an intervention's causal impact, and (2) allow for design-based inference, meaning that the physical act of randomization largely justifies the statistical assumptions made. However, RCT sample sizes are often small, leading to low precision; in many cases RCT estimates may be too imprecise to guide policy or inform science. Observational studies, by contrast, have strengths and weaknesses complementary to those of RCTs. Observational studies typically offer much larger sample sizes, but may suffer confounding. In many contexts, experimental and observational data exist side by side, and recent work has sought to integrate "big observational data" with "small but high-quality experimental data" to get the best of both worlds. Such approaches hold particular promise in the field of education, where RCT sample sizes are often small due to cost constraints, but automatic collection of observational data, such as in computerized educational technology applications, or in state longitudinal data systems (SLDS) with administrative data on hundreds of thousands of students, has made rich, high-dimensional observational data widely available. In this talk I will discuss a framework that allows researchers to employ machine learning algorithms to learn from the observational data, and use the resulting models to improve precision in randomized experiments. Importantly, there is no requirement that the machine learning models are "correct" in any sense, and the final experimental results are guaranteed to be exactly unbiased. Thus, there is no danger of confounding biases in the observational data "leaking" into the experiment.

TITLE: Improving
the Power of
Education
Experiments with
Auxiliary Data

Speaker:

**Johann Gagnon-
Bartsch**

Assistant Professor
Department of Stat
University of Michigan

Time & Place:

Wednesday,

October 16, 2019

4:00, Room 140

Bardeen

Cookies & Coffee @

3:30, Rm 1210 MSC

