Opportunities and Insights (and Some Accessible R Packages) for the Planning and Analysis of Multisite and Cluster Randomized Trials

Instructors: Luke Miratrix, Harvard University, and Michael Weiss, MDRC

Description: In this interactive workshop we will discuss the conceptualization, planning, and analysis of different types of randomized trials found in educational effectiveness research. The workshop will be an interactive lecture, with didactic portions intermixed with participant discussion of a set of provided conceptual questions and some in-class coding with sample data.

We first discuss the myriad models one might use to estimate average treatment effects, and shine light on nuances of how they differ in terms of how they perform and what they are actually targeting, if there is impact variation. We then give a brief overview on conducting easy power analyses, focusing on how different choices of a planned analysis can impact power. We give a quick introduction to a package that allows easy use of any of the methods discussed, and turn to methods for estimating the degree of cross-unit variation of impacts. Time permitting, we will also touch on other forms of estimating impact heterogeneity.

Our session will have the following major portions:

I: Estimands and design

Here we discuss different ways of conceiving of an average impact of a treatment in multisite or multilevel trial contexts. We also compare and contrast multisite trials vs. cluster randomized trials, and discuss the tradeoffs of these types of design. We focus on how evaluators could focus on an evaluation sample or broader population, and on how they may focus on the average impact for the individual vs. for site or other grouping of units.

As part of this section we will have a turn and talk: we will present three scenarios and have people debate in small groups which are the most appropriate estimands. We will then gather all together and talk about pros and cons of different choices made.

II: The estimators

In this more didactic portion of the seminar, we will walk through the different estimators and flag important aspects of each. This may include: what estimand does each estimator most naturally target? Is the estimator potentially biased with respect to that estimand?

We will also review some results showing to what degree estimates will differ depending on these design decisions. We will in particular draw attention to differences in the quality

of uncertainty estimation and various bias-variance tradeoffs across the estimation strategies.

III: Planning and power

We will show how to explore power, MDES, and needed sample sizes using the PUMP package. We will highlight how power changes depending on the estimand or estimator selected.

In a short exercise, participants will explore a specific scenario to answer the question of how much the sample size would have to go up if we were targeting a superpopulation estimand. We will discuss what that means in terms of planning.

IV: And impact variation?

We will discuss different forms of impact variation, and how it might be estimated in different contexts. We compare cluster randomized to multisite experiments in terms of the ability to assess impact heterogeneity.

We then will teach the currently used approaches for estimating cross-site impact variation in multisite contexts using the FIRC model and meta-analytic techniques. We will also review how much variation was found in recent work in this area (Weiss et al. 2017, Bloom et al. 2017). We finally demonstrate estimation using a simple R package, blkvar. Participants will then take a given dataset and estimate the degree of cross-site impact heterogeneity using the package.