A Statistical Framework and Tools for Planning Multilevel Randomized Cost-Effectiveness Trials

In recent years, there has been increasing demand for rigorous evidence on both the impacts of educational interventions and the net cost of producing those impacts. In response to the requirements for cost-effectiveness analysis (CEA) from major funding agencies, some educational interventions have started to evaluate the cost-effectiveness (or cost-benefit) of their interventions (e.g., Jacob et al., 2016; Mustafa, 2018; Steele et al., 2018; Unlu et al., 2015). We anticipate that increasing proportions of education experiments will incorporate cost and cost-effectiveness study components. It is now widely accepted among the education research community that statistical power is an important consideration in the design of impact evaluations. However, most commonly, the evaluators generally design their studies to meet established goals concerning the statistical power of the focal outcomes (e.g., standardized test scores, grade progression rates, high school graduation rates). That is, they tend to worry if the sample size is adequate to achieve a desired probability of detecting treatment effects of a magnitude that have policy relevance if they exist (Boruch & Gomez, 1977; Cohen, 1988). Education evaluators often ignore costs altogether and, when they do include a cost analysis component in their study design, they rarely consider the statistical power of the estimated cost-effectiveness measures in the study design phase (Belfield & Bowden, 2019). One major challenge of incorporating CEA into impact evaluations is covering the cost of data collection. As a result, evaluators sometimes choose to restrict their cost studies to a subsample of the participating schools, raising concerns about both power and representativeness.

In an ideal world, education evaluations would be designed to produce good estimates of both the impacts on the focal outcomes of interests and the costs of producing those outcomes. This means it is important to consider threats of bias and statistical power in cost estimates as well as in estimates of impacts on the focal outcomes. There are well-articulated methods for conducting power analysis for measures of the effectiveness of the treatment in RCTs (e.g., Dong & Maynard 2013; Spybrook et al., 2011). However, there is not a well-developed literature laying out how to conduct power analysis for RCTs that focus on the evaluation of the cost-effectiveness of the treatment, which are commonly referred to as randomized cost-effectiveness trials (RCETs). RCET power analyses need to incorporate the estimated variances of both the effectiveness measures and the cost measures, to ensure that the sample design offers a “good enough” chance (e.g., *power* ≥ 0.80) to detect meaningful size differences in the cost-effectiveness of the treatment versus the control.

Recent work in health science (Manju, Candel, & Berger, 2014, 2015; Manju, Candel, & van Breukelen, 2019) provides formulas and an interactive computer program to calculate power for two-level randomized cost-effectiveness trials, where the treatment is either at the first level (e.g., patients) or the second level (e.g., health care providers). The counterpart example in education would be treatment at the student or school level. However, those formulas and tools cannot be directly applied to educational studies for several reasons. First, the power analysis methods in health science assume that individual level data for the cost of implementing the treatment are available, while such data are commonly missing or very difficult to collect in for educational evaluations (Levin & Belfield, 2015). Second, these methods developed for health care evaluations do not account for the effects of covariates, which are commonly available and used in educational evaluations to improve precision of impact estimates (e.g., Bloom et al., 2007; Hedges & Hedberg, 2007). Third, prior approaches only considered two-level designs (e.g., students nested within classrooms), while educational interventions usually have more complicated nesting structure (e.g., students nested within classroom, and classrooms nested within schools). Fourth, the current power analysis tool (e.g., Manju et al., 2019) is mainly oriented toward evaluations in the health sciences, and thus does not fit the need of educational researchers very well to plan other types of study designs (e.g., three-level designs) common in education.

A recent attempt in education is Li, Dong, & Maynard (2020), which considers two-level multisite randomized cost-effectiveness trials where treatment is at level 1. However, a general statistical framework and tool for planning two- and three-level cost-effectiveness trials, for which treatment is assigned at different level and cost information is available at different level, still lack. This project is designed to address these limitations and contributes to the design and analysis of educational MRCETs to accommodate both two- and three-level models, various assumptions about costs and covariates, and alternative sample designs and analytic modelling assumptions. Specifically, we will develop new formulas for estimating the statistical power to MRCETs with 10 alternative designs and implement these formulas to an accessible and user-friendly software program that is designed to fit educational researchers’ needs.