

SREE 2020 Conference
“In the Pipeline” Poster Session for Works in Progress
Daniel Litwok¹, Azim Shivji¹, & Rob Olsen²

Title: Selecting Districts and Schools for Impact Studies in Education: A Simulation Study of Different Methods

Background/Context: In most randomized trials in education, districts are not required to participate, and incentives to participate can be weak. Furthermore, even when districts do participate, schools may have the choice to opt out. As a result, researchers often have to recruit many more schools than they need to reach their sample size target. This project uses the Common Core of Data (CCD) to simulate several strategies for site selection and estimate the external validity bias when site participation is optional.

Purpose/Objective/Research Question: The study documents the performance of methods of site selection in terms of external validity. The study tests established methods for site selection: stratified random site selection (e.g., Olsen & Orr, 2016) and methods for systematically selecting sites to match the target population (e.g., Tipton, 2013). The study also develops and tests methods for selecting replacement sites modelled after methods used in survey research. These methods are compared to purposive site selection—“business as usual” in randomized trials (Olsen et al., 2013)—which can produce biased impact estimates (e.g., Bell et al., 2016). We simulate purposive selection rules based on factors that researchers often consider when selecting sites.

Setting/Population/Participants/Subjects: Samples of sites are selected from all regular K-5 public schools in the United States (as downloaded from the CCD) for a hypothetical intervention.

Intervention/Program/Practice: The study simulates an evaluation of a hypothetical intervention in a K-5 public school. The intervention’s impact is a function of several characteristics at the school level (size and percent receiving free or reduced price lunch) and at the district level (size, expenditures per pupil, and region). We also vary several parameters across simulations to create a variety of scenarios that mimic actual evaluations: the sample size, the extent of variation in impacts across sites (Weiss et al., 2017), the proportion of variation explained by covariates, and the proportion of chosen districts that agree to participate in the evaluation.

Research Design: The study simulates the impact of the intervention for all schools within the population. Then, the study “recruits” sites using four different site selection approaches: purposive sampling, random sampling, systematic sampling, and a hybrid approach (where schools are selected randomly but replaced systematically). Under each of these approaches we select a group of districts and simulate their participation; then we select schools within districts that participate and simulate their participation. We repeat this process until the target number of schools for the intervention is reached. Based on the sites that agree to participate in the

¹ Abt Associates

² Westat

hypothetical evaluation, the study can estimate the impact of the intervention for that particular sample using standard methods for evaluation of a randomized trial (e.g., a hierarchical model).

Analysis: In total, for each model we select 500 purposive samples, 500 random samples, 500 systematic samples, and 500 hybrid samples. We estimate external validity bias by comparing the average impact from the sites simulated to participate in the study to the average impact from all sites in the population, conduct a formal test for the presence of external validity bias, and estimate the mean squared error for each site selection method.

Findings/Results: As an “In the Pipeline” proposal, there are no findings to report in this proposal. Our work in building the simulation study is underway, and we plan to summarize our progress in the poster session. We look forward to hearing the input of our SREE colleagues in the poster session as we make progress toward generating findings and interpreting results.

References:

Bell, S. H., Olsen, R. B., Orr, L. L., & Stuart, E. A. (2016). Estimates of external validity bias when impact evaluations select sites nonrandomly. *Educational Evaluation and Policy Analysis*, 38(2), 318-335.

Olsen, R. B., & Orr, L. L. (2016). On the “where” of social experiments: Selecting more representative samples to inform policy. *New Directions for Evaluation*, 2016(152), 61-71.

Olsen, R. B., Orr, L. L., Bell, S. H., & Stuart, E. A. (2013). External validity in policy evaluations that choose sites purposively. *Journal of Policy Analysis and Management*, 32(1), 107-121.

Tipton, E. (2013). Stratified sampling using cluster analysis: A sample selection strategy for improved generalizations from experiments. *Evaluation review*, 37(2), 109-139.

Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil, A. E., & Cullinan, D. N. (2017). How much do the effects of education and training programs vary across sites? Evidence from past multisite randomized trials. *Journal of Research on Educational Effectiveness*, 10(4), 843-876.