

SREE 2020 Conference
“In the Pipeline” Poster Session for Works in Progress
Emily Diaz, Stephen Bell, & Rob Olsen

Title: Aiming for Representative Samples in Impact Studies: Evidence from Head Start

Background/Context: Most randomized trials in education do not choose sites to be formally representative of any broader population (Olsen, Orr, Bell & Stuart, 2013). This is probably because sites are not typically required to participate in these studies—and often choose not to participate when selected. One of the few exceptions is the Head Start Impact Study, which was conducted by Westat in a probability sample of 84 grantees and 383 Head Start centers (Puma et al., 2010). Conducting the study in a representative sample of Head Start programs was made possible by a Congressional mandate, active support for the study by the Department of Health and Human Services, and the success of the project team in negotiating with Head Start programs.

Because the Head Start Impact Study successfully recruited a nationally representative sample of grantees and centers, it was able to produce national estimates of the program impact. However, this is probably not feasible in most randomized trials where participation incentives are weak. Hence, the Head Start Impact Study data offers not only an opportunity for reanalysis to estimate how much bias would be introduced when a large share of selected programs opt out of the study, but also how that bias can be minimized by different strategies for selecting sites and replacement sites for selected sites which opt out of the study.

Purpose/Objective/Research Question: The purpose of the study is to test different methods for selecting sites—and replacement sites—in randomized trials, particularly in those likely to face low study participation rates by selected sites. In addition, the study aims to test:

- different methods for selecting replacement sites, including methods drawn from survey research, and
- different combinations of methods for selecting sites initially and selecting replacement sites when needed.

Setting/Population/Participants/Subjects: The Head Start Impact Study was conducted in 84 program grantees and 383 centers. The sample included 2,559 three-year-old children and 2,108 four-year-old children for a total of 4,667 children.

Intervention/Program/Practice: Head Start provides free center-based care and early childhood education to children from low-income families. The goal of Head Start is to prepare children for school and teach their parents child development skills. Head Start was founded on a “whole child” model to provide a wide range of supports to children and their families, including education, medical, dental, nutrition, and mental health (Puma et al., 2010).

Research Design: The Head Start Impact Study was designed as a randomized trial. Our reanalysis of the study data is designed as a simulation study based on the framework of the actual study data. To create a pseudo population, we will sample with replacement from the Head Start centers that participated in the original study. Then we will simulate 500 samples for

each of several combinations of methods of selecting sites and selecting replacement sites. In particular, we plan to test three possible approaches to selecting sites:

- (1) purposive site selection that favors sites with relatively high probabilities of agreeing to participate, which is standard practice in randomized trials;
- (2) random site selection with known probabilities, as used in the Head Start Impact Study and many large-scale surveys, and recommended in Olsen and Orr (2016); and
- (3) systematic site selection to match the population, as developed in Tipton (2014) and used in Tipton & Matlen (2019).

In addition, we will test two different ways of selecting replacement sites:

- (1) the same method used to select sites originally and
- (2) a method that selects the most similar site to the sites being replaced.

To select samples of sites for the analysis, we will build models of purposive site selection, random site selection, systematic site selection, and site self-selection. For the model of purposive site selection, the probability of selecting a center will be positively associated with size, dosage, treatment contrast, and demand for services—all factors that would plausibly affect researchers' decisions to recruit a Head Start center if centers were not expected to participate. The model of systematic site selection that we will test, from Tipton (2014), is completely deterministic. However, the model of site self-selection into the study will be probabilistic regardless of how sites are selected.

Analysis: First, we will estimate the Population Average Treatment Effect (PATE). Second, for each method of selecting and replacing sites we will generate 500 samples and estimate the Sample Average Treatment Effect (SATE) for each sample. To estimate the bias for each method, we will calculate the average SATE and subtract the PATE; to estimate the variance for each method, we will estimate the variance of the SATE estimates across the 500 samples. Then we will calculate the Mean Squared Error from the estimates of bias and variance.

Findings/Results: Our work in building the simulation study is underway, and we plan to summarize our progress in the poster session. We look forward to input from our SREE colleagues as we work to complete the study and publish the results.

References:

- 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.
- Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ..., Spier, E. (2010). *Head Start Impact Study: Final Report*. Rockville, MD: Westat.

Tipton, E. (2014). Stratified sampling using cluster analysis: A sample selection strategy for improved generalizations from experiments. *Evaluation review*, 37(2), 109-139.

Tipton, E., & Matlen, B. J. (2019). Improved Generalizability Through Improved Recruitment: Lessons Learned From a Large-Scale Randomized Trial. *American Journal of Evaluation* 40(3), 414–430. <https://doi.org/10.1177/1098214018810519>.