'generalize': Statistical Software for Implementing Methods to Generalize Randomized Trial Findings to a Well-Defined Target Population

Benjamin Ackerman, Ian Schmid, Trang Quynh Nguyen, Marissa Seamans & Elizabeth A. Stuart, Johns Hopkins Bloomberg School of Public Health

Background: Randomized controlled trials (RCTs) are considered the gold standard for estimating the average effect of a treatment in a study population. Evidence from trials are often used by researchers in the education field to examine the effectiveness of educational interventions. While trials have strong internal validity by design, there is growing concern over the potentially poor external validity, or generalizability, of trial results to a target population of interest (Bell et al., 2016; Cook, 2014; Tipton, 2014). In particular, the sample average treatment effect estimated in the trial (SATE) will be a biased estimate of the target population average treatment effect (TATE; Olsen et al., 2013) if there are treatment effect moderators whose distributions differ between the trial and target population. Statistical methods currently exist to obtain a more accurate estimate of the TATE that accounts for differences between the trial and target populations; however, little practical guidance on acquiring the necessary population data and on implementing these methods is available for applied researchers interested in utilizing the generalizability methods.

Purpose: We provide an overview of post-trial statistical methods for assessing and improving upon the generalizability of a randomized trial to a well-defined target population. We then provide software to implement the methods in R using the "generalize" package, and discuss several practical considerations for researchers who wish to utilize these tools. These include the importance of acquiring population-level data to represent the target population of interest, and the challenges of data preprocessing and harmonization between trial and population data. Finally, we illustrate the implementation of the "generalize" using a randomized trial related to methamphetamine dependence and a target population of substance abuse treatment seekers.

Methods: The "generalize" package contains two core functions: *assess* and *generalize*. The *assess* function evaluates differences between the trial sample and the target population based on a specified list of common covariates. This is done in several ways: first, *assess* generates a table summarizing the means of each characteristic in the trial and population, along with an absolute standardized mean difference (ASMD) between the two. Then, *assess* implements propensity score-type methods to estimate the probability of trial participation based on the observed pre-treatment covariates. The distributions of trial participation probabilities in the trial and population are compared to examine how different the trial may be from the population, which can help researchers determine whether it is feasible to generalize findings from that trial to the specified target population. The *generalize* function enables users to estimate the TATE using one of the following three broad classes of methods: one set based on using the probability of trial participation to equate the trial sample and population, one set based on flexible outcome models used to predict outcomes in the population, and a third that combines both together.

Conclusion: When recruiting fully representative samples or altering study design to strengthen external validity is infeasible, statistical methods for estimating target population effects are helpful tools that allow education researchers to better estimate population average treatment

effects post-hoc. In this poster, we provide statistical software that enables researchers to implement these methods, and offer guidance in acquiring and preparing the necessary data to estimate population effects from randomized trials.

References:

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

Cook, T. (2014). Generalizing causal knowledge in the policy sciences: External validity as a task of both multi-attribute representation and multi-attribute extrapolation. *Journal of Policy Analysis and Management* 527-536. DOI: 10.1002/pam.

Olsen, R., Bell, S., Orr, L., and Stuart, E.A. (2013). External Validity in Policy Evaluations that Choose Sites Purposively. *Journal of Policy Analysis and Management* 32(1): 107-121. NIHMS 382967

Tipton, E. (2014). How generalizable is your experiment? Comparing a sample and population through a generalizability index. *Journal of Educational and Behavioral Statistics, 39*(6): 478 – 501.