Sensitivity Analysis for an Unobserved Moderator in Trial-to-Target-Population Generalization of Treatment Effects

Benjamin Ackerman, Trang Quynh Nguyen & 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. Researchers in the education field often use evidence from RCTs to examine the effectiveness of educational interventions. While trials have strong internal validity by design, there is growing concern over potential poor external validity, or generalizability, of trial results to a target population of interest (Bell et al., 2016; Cook, 2014; Tipton, 2014). In particular, if there are treatment effect moderators whose distributions differ between the trial and target population, then 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). Statistical methods currently exist to obtain a more accurate estimate of the TATE that accounts for differences between the trial and target populations; however, these methods assume that all effect moderators are observed in both the trial and the dataset representing the target population. In practice, once data are identified to represent a target population, researchers often discover that some effect moderators relevant to and collected in the trial are not present in the target population data (Stuart and Rhodes, 2017). Researchers may be worried about how unobserved moderators impact the ability to estimate population treatment effects from RCT samples.

Purpose: In this talk, we present sensitivity analyses to estimate the TATE when: 1) an effect moderator is observed in a trial, but not in the target population, and 2) an effect moderator is unobserved in both the trial and the target population. The methods are described in Nguyen et al., 2017; the talk will describe the methods in a broadly accessible way and discuss implications for education research.

Methods: In the case of a partially unobserved moderator, we describe and implement methods to estimate the TATE based on an outcome model, on full weighting adjustment, and on an outcome model combined with partial weighting. Plausible ranges are specified wherever there are unknown values due to unobserved moderation, allowing us to obtain a range of values for the TATE, along with a confidence band in some methods. In the case of a fully unobserved moderator, we describe methods for TATE estimation based on both an unweighted and a weighted bias formula.

Simulation and Data Application: The performance of the sensitivity analysis methods for the partially unobserved moderator are tested through simulation. We vary levels of model misspecification, the correlation between the fully observed and partially unobserved moderators, and whether the fully and partially observed moderators are binary or continuous. We also apply the methods to a smoking cessation RCT for drug and/or alcohol-dependent adults, and generalize the results to respondents of the National Survey on Drug Use and Health (NSDUH). We consider baseline cigarette addiction severity as the partially unobserved moderator, since it moderates treatment effects and is measured in the trial but not in NSDUH.

Results: Under model misspecification with respect to the observed moderator, the combined weighted-outcome-model based method is less biased than the outcome-model based method. When the model is misspecified with respect to the unobserved moderator, and when there is a positive correlation between the observed and unobserved moderators, the weighted-outcome-model based method is also less biased. The data example depicts how a range of values for the mean addiction score in the target population produces a range in estimating the TATE. While both the outcome-model based method and the weighted-outcome-model based method produce similar point estimates, the weighted-outcome-model based method yields slightly wider confidence intervals due to the weighting.

Conclusion: In this paper, we present approaches for researchers and policymakers to assess how sensitive generalized RCT findings are to potentially unobserved moderators. They allow researchers to implement relatively straightforward methods to estimate the TATE and to better utilize RCT findings in their decision making, even when the trial or the target population data lack certain important characteristics. Given that observing all effect moderators is almost never possible the methods have potential wide applicability across education research and decision making when aiming to estimate population effects from randomized trial data.

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.

Nguyen, T.Q., Cole, S., Ebnesajjad, C., and Stuart, E.A. (2017). Sensitivity Analysis for an Unobserved Moderator in RCT-to-Target-Population Generalization of Treatment Effects. *The Annals of Applied Statistics* 11(1): 225-247.

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

Stuart, E.A., and Rhodes, A. (2017). Generalizing Treatment Effect Estimates from Sample to Population: A case study in the difficulties of finding sufficient data. Forthcoming in *Evaluation Review* 41(4): 357-388

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.