Abstract Title Page

Not included in page count.

Title: Student Mobility, Dosage, and Principal Stratification in Clustered RCTs of Education Interventions

Authors and Affiliations: Peter Z. Schochet, Senior Fellow, Mathematica Policy Research, Inc.

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

> In RCTs of educational interventions, random assignment is often performed at the school (or classroom) level rather than at the student level. Under these designs, data are often collected on repeated cross-sections of students as they progress into higher grades. Thus, because students can move into and out of the study schools during the follow-up period, a critical issue for these RCTs is student mobility, which can affect the students for whom data are collected, the analysis samples, the choice of impact estimators, and the interpretation of study findings.

Student mobility is common, especially among low-income, elementary and middle school students who are often the focus of large-scale RCTs. Only about 55 percent of America's kindergartners remain in the same school by the end of third grade (Burkam et al., 2009), and the figures are considerably lower for African American children, students in poverty, and students in urban centers who often move both within and across school districts (de la Torre & Gwynne, 2009; Hanushek, Kain, & Rivkin, 2004; Pianta & Early, 2001; Sampson & Sharkey, 2008).

Research has found that the characteristics of students who change schools tend to differ from those who do not, and that student mobility can have a negative effect on student learning (Raudenbush et al., 2010; Reynolds et al., 2009). Thus, in school-based RCTs, the composition of students in the study schools could change over time, which could influence the size of the impact estimates and their interpretation. These issues are more problematic if the intervention has an effect on student mobility (Bloom, 2004).

There are several causal treatment effect parameters for school-based RCT designs that address student mobility. The intention-to-treat (*ITT*) parameter pertains to the average treatment effect (*ATE*) on key study outcomes for all treatment and control group students *in the study schools at baseline*. Under the *ITT* design, the initial cohort of students is followed over time, including those who remain in their initial study schools or transfer to other study schools—"*stayers*"— and those who leave the study schools— "*leavers*". This approach yields causal impact findings that pertain to a well-defined student population. However, the *ITT* parameter may be difficult to interpret, because students in the *ITT* sample may have different levels of exposure to the intervention, which may not occur if the intervention were adopted more broadly. Furthermore, the *ITT* design is rarely adopted for clustered education RCTs because of the high cost of collecting outcome data for leavers across a potentially large number of geographically-dispersed school districts.

Instead, a much more common approach is to focus on the place-based (*PB*) impact parameter, which pertains to the *ATE* for all students who are enrolled in the treatment and control schools at the follow-up data collection point (Boruch & Foley 2000). Under the *PB* design, the analysis sample includes stayers as well as students who transfer into the study schools between data collection points—"*new entrants*".

The *PB* approach yields unbiased, school-based causal effect estimates under real-world conditions that include student mobility. However, if student mobility is common and is influenced by the intervention, the composition of students in the treatment and control schools could differ at each follow-up point. In this case, the *PB* parameter could confound intervention effects on student mobility with intervention effects on student achievement. Isolating these latter achievement effects, however, is likely to be of more interest to policymakers, because intervention effects on student mobility would not be germane if the intervention were adopted more broadly. Furthermore, similar to the *ITT* parameter, the *PB* parameter pertains to students with different levels of exposure to the intervention, which complicates the interpretation of study findings. Finally, baseline pretest data are often missing for new entrants, which can further complicate the analysis.

Purpose / Objective / Research Question / Focus of Study:
Description of the focus of the research.

> To address these issues, this article introduces an alternative impact parameter for group-based RCTs with student mobility—the survivor average causal effect (*SACE*)—that pertains to the subpopulation of original cohort students who would remain in their baseline study schools in either the treatment or control condition. The *SACE* parameter has a clear interpretation, because it pertains to students who would receive maximum exposure to the intervention in the treatment condition, and who would receive a common array of intervention services within each treatment school. For many education RCTs, intervention exposure will be strongly correlated with intervention *dosage*, and thus, the *SACE* parameter is likely to be highly relevant for dosage analyses. Consequently, the methods presented in this article can be used to address the following research question that is often of interest to education policymakers and evaluators: What are intervention effects for the subgroup of students who receive full exposure (or the full dose) of intervention services?

This research question is germane to the SREE conference theme, because the *SACE* parameter provides policy-relevant information on the heterogeneity of treatment effects for the subgroup of the full study population who remain in the study schools throughout the follow-up period. Clearly, levels of intervention dosage cannot always be mandated. However, understanding program effects for those who are exposed to intervention services for the full follow-up period provides policy-relevant information on likely program effects that would be observed if intervention implementation was improved or became more widespread. Thus, the estimation of the *SACE* parameter can strengthen overall RCT findings.

Setting: NA
Population / Participants / Subjects: NA
Intervention / Program / Practice: NA

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.

> The primary contribution of the article is the development of rigorous statistical methods to estimate dosage effects in clustered education RCTs with student mobility, by systematically examining the identification and estimation of the *SACE* parameter in this context. As discussed

further below, the statistical framework used in the article relies on a principal stratification (PS) and potential outcomes approach, where students are allocated to principal strata based on their potential mobility decisions during the follow-up period. Importantly, identification of the *SACE* parameter is driven by distributional assumptions about potential student outcomes within each principal stratum. Thus, this approach for conducting an RCT dosage analysis does not rely on (1) hard-to-find instrumental variables to identify causal effects for analyses that compare the outcomes of self-selected samples of stayers in the treatment and control groups (Heckman & Robb 1985), or (2) propensity score matching methods (Rosenbaum & Rubin, 1983; Schochet & Burghardt, 2007) that rely on highly predictive covariates that are correlated with study outcomes to match treatment stayers with comparable controls (or vice versa).

The article is also the first to use the PS approach to define and contrast the *PB*, *ITT*, and *SACE* impact parameters, and discusses simple contextual analyses for comparing and interpreting these parameters. The article also defines a new impact parameter—the new entrants average causal effect (*NACE*) parameter—that pertains to new entrants who would enter the study schools in the presence or absence of the evaluation.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

> The article uses a principal stratification approach following the work of Frangakis & Rubin, (2002), Little & Yau (1998), Rubin (2006), and Zhang, Rubin, & Mealli (2009). In the present context, the PS approach for the *SACE* parameter involves allocating students in the initial cohort sample to unobserved principal strata defined by students' *potential* mobility statuses in the treatment and control conditions. Under this approach, students can stay in their study schools during the follow-up period, transfer to other treatment schools, transfer to other control schools, or leave the study school districts. There are many possible principal strata because a student can transfer schools at any time during the follow-up period. Note that without further restrictions, it is not possible to determine the specific principal stratum for any student, because mobility decisions are only observed in either the treatment or control condition, but not both.

Nonetheless, the *SACE* parameter can be estimated using ML methods for finite mixture models, the EM algorithm (Dempster et al. 1977), and a bootstrap approach for estimating standard errors of the impact estimates to account for school-level clustering. Identification of the *SACE* parameter is driven by distributional assumptions about potential student outcomes within each principal stratum.

The article discusses this statistical approach, including the model specifications and assumptions, the likelihood functions, and the estimation of the parameters and their variances.

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.

> The article demonstrates the developed PS approach using data from the Social and Character Development (SACD) Research Program, a large-scale, school-based RCT funded by the Institute of Education Sciences (IES) and the Centers for Disease Control and Prevention (CDC) (SACD Research Consortium, 2010). The SACD study evaluated promising school-based

interventions in 84 schools in seven sites that were designed to promote positive social and character development among elementary school children.

For this study, student mobility was common over the three-year follow-up period. Only about two-thirds of the students in the Spring 2007 follow-up interview sample were original cohort stayers. Thus, although the SACD evaluation study found no statistically significant *PB* impact estimates on key study outcomes, it is possible that the SACD interventions may have had a beneficial *SACE* effect for the two-thirds of original cohort students who received maximum exposure to intervention services.

Research Design: NA
Data Collection and Analysis: NA

Findings / Results:
Description of the main findings with specific details.
(May not be applicable for Methods submissions)

> For the SACD case study, very small treatment effects on student mobility were found during the follow-up period, but original cohort stayers in both research groups were found to be less disadvantaged than both the leavers and new entrants, highlighting the important point that the *SACE* parameter in education RCTs will often pertain to a different student population than for the *PB* impact parameter. Most importantly, we found that the *PB*, *SACE*, and *NACE* impacts on three key study behavioral and academic outcomes were all statistically insignificant, suggesting that the SACD interventions did not have beneficial effects, even for the subpopulation of stayers who received maximum exposure to intervention services.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

> The primary contribution of this article is the development of rigorous statistical methods to estimate dosage effects in education RCTs. The estimation of the *SACE* parameter for a dosage analysis provides a viable alternative to instrumental variable and propensity scoring methods. The estimation of the SACE parameter can strengthen the overall findings of education RCTs, by providing policy-relevant information on treatment effects for maximum-exposure students.

The case study and estimation methods in this article apply to a design with two mobility statuses (stayers and leavers) and three principal strata. The same methods can be adapted to designs with additional mobility groups and principal strata. However, in such designs, computational complexities (such as model convergence) could arise as the number of model parameters significantly increase. An area for future research is to empirically assess the computational feasibility of using the principal stratification approach for designs with multiple stayer and leaver groups, and the specification of additional assumptions (such as monotonicity conditions) that could be required to reduce the number of principal strata and model parameters.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Bloom, H. (2004). Randomizing groups to evaluate place-based programs. MDRC Working Paper: New York, New York.

Boruch, R. G., & E. Foley. 2000. The honestly experimental society: sites and other entities as the units of allocation and analysis in randomized trials." In *Validity and Social Experimentation: Donald Campbell's Legacy (Volume 1),* edited by Leonard Bickman. Thousand Oaks, CA: Sage Publications.

Burkham, D.T., V. Lee, & J. Dwyer (2009). School mobility in the early elementary grades: frequency and impact from nationally representative data. Paper commissioned by the National Academy of Sciences Committee on the Impact and Change in the Lives of Young Children, Schools, and Neighborhoods. Washington, D.C.

De La Torre, M. & J. Gwynne (2009). Changing schools: a look at student mobility trends in Chicago public schools since 1995. Consortium on Chicago School Research: Chicago.

Dempster, A., N. M. Laird, & D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B. 39, 1-38.

Frangakis, C. E. & D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics*, 58, 20-29.

Hanushek, E., J. Kain, & S. Rivkin (2004). Disruption versus Tiebout improvement: the costs and benefits of changing schools. *Journal of Public Economics*, 88, 1721-1746.

Heckman, J. J. & R. Robb Jr. (1985). Alternative methods for evaluating the impact of evaluations. *Journal of Econometrics,* 30, 239-267.

Little, R. J. & L. Yau (1998). Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychological Methods*, 3, 147-159.

McLachlan, G. & D. Peel (2002). *Finite mixture models.* John Wiley and Sons, New York, NY.

Pianta, R.C., & D. Early (2001). Turnover in kindergarten classroom membership in a national sample. *Early Education and Development,* 12, 239-252.

Raudenbush, S. W., J. Marshall, & E. Art. (2010) Year-by-year cumulative impacts of attending a high-mobility elementary school on children's mathematics achievement in Chicago, 1995 – 2005. Working Paper: U. of Chicago Department of Sociology.

Reynolds, A. J., C. Chen & J. E. Hebers (2009). School mobility and educational success: a research synthesis and evidence on prevention. Paper commissioned by the National Academy of Sciences Committee on the Impact and Change in the Lives of Young Children, Schools, and Neighborhoods. Washington, D.C.

Rosenbaum, P. & D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects." *Biometrika,* 70. 41-55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Education Psychology*, 66, 688-701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate, *Journal of Education Statistics*, 2(1), 1-26.

Rubin, D. B. (2006). Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Statistical Science.* 21, 299-309.

SACD Research Consortium (2010). Efficacy of schoolwide programs to promote social and character development and reduce problem behavior in elementary school children. Final Report: Institute for Education Sciences, U.S. Department of Education, Washington D.C.

Sampson, R. J. & P. Sharkey (2008). Neighborhood selection and the social reproduction of concentrated racial inequality. *Demography* 45, 1-29.

Schochet, P. Z. & J. Burghardt (2007). Using propensity scoring to estimate program-related subgroup impacts in experimental program evaluations. *Evaluation Review,* 31, 95-120.

Zhang, J.. D. B. Rubin, & F. Mealli (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification, *Journal of the American Statistical Association,* 104, 166-176.

Appendix B. Tables and Figures
Not included in page count.