Leveraging Differential Sampling Cost Structures to Improve the Design of Cluster-Randomized Trials

Zuchao Shen, University of Florida

&

Ben Kelcey, University of Cincinnati

Purpose

A common design strategy is to identify a sampling scheme that maximizes power under a predicted cost structure and fixed budget (i.e., optimal design). Conventional frameworks have largely assumed equal costs of sampling treatment and control units (e.g., Raudenbush, 1997; Konstantopoulos, 2009, 2011; Hedges & Borenstein, 2014). The purpose of this study was to develop a more flexible framework that allows for and leverages differential costs to improve design efficiency.

Background

Raudenbush (1997) developed an optimal design framework that identifies the sample allocation that produces the minimum variance of a treatment effect under a fixed budget. Specifically, in a two-level CRT with a total number of J clusters and n individuals in each cluster, let the cost of sampling an additional individual and cluster be C_1 and C_2, respectively such that the total cost of a study is $m = J(C_1 n + C_2)$. Under this budget constraint, the optimal sample allocation is

$$n = \sqrt{\frac{(1-\rho)(1-R_1^2)}{\rho(1-R_2^2)}} \sqrt{\frac{C_2}{C_1}}\,, \tag{1}$$

where ρ is the unconditional intraclass correlation coefficient, R_1^2 and R_2^2 are the proportions of outcome variance explained by covariates at the individual and cluster level, respectively. Once an optimal individual-level sample size, n, is identified, the corresponding level-two sample size, J, is determined by the desired power level or a given budget.

Implicit in this framework is that the costs of sampling a unit in the treatment condition are equal to those in the control condition. More generally, virtually every subsequent development of the optimal sampling frame work has adopted this assumption (e.g., multisite trials in Raudenbush & Liu, 2000; three- or four-level designs in Konstantopoulos, 2009, 2011 and Hedges & Borenstein, 2014). In many practical situations, however, the sampling cost of units may additionally vary by treatment condition as well as levels of hierarchy (e.g., Liu, 2003; Mosteller, 1995; Springer et al., 2011). In a simple recent example, in a cluster-randomized evaluation of whether incentives in teacher performance improve student outcomes (Springer et al., 2011), teachers in the experimental group were eligible to receive a bonus payment of up to $15,000 per year based on their students' performance in tests whereas teachers in the control condition were not offered any incentive. As a result, the costs of sampling each additional teacher in the experimental group typically exceeded the cost associated with sampling an additional control teacher.

Differential sampling costs among treatment conditions have also been documented in studies evaluating the effects of class size (Mosteller, 1995), community health interventions (e.g., Glynn et al., 1995), training and professional development (e.g., Hiscock et al., 2008; Greenleaf et al., 2011; Jacob, Goddard, Kim, Miller, & Goddard, 2015; Jayanthi, Gersten, Taylor, Smolkowski, & Dimino, 2017).

Liu (2003) partially relaxed these cost assumptions by allowing cost differences between treatment conditions. However, the Liu (2003) framework presents a type of expanded but still constrained optimal design because it fixes the individual-level sample size. Next, we present a more flexible framework for two-level CRTs.

Optimal Design of Two-Level CRTs With Unequal Costs

Our framework begins by first differentiating the costs between the treatment and control conditions. We assign c_1 and c_1^T as the cost of enrolling each additional individual in the control and treatment conditions, respectively, and we use c_2 and c_2^T as the respective cost of sampling each additional control and treatment cluster. The resulting total cost of a study becomes $m = (1-p)J(c_1 n + c_2) + pJ(c_1^T n + c_2^T)$. We can then specify the number of clusters as a function of the remaining parameters as

$$J = \frac{m}{(1-p)(c_1 n + c_2) + p(c_1^T n + c_2^T)}. \tag{2}$$

Substituting this equation into the error variance of the treatment effect gives

$$\sigma_\delta^2 = \frac{n\rho(1-R_2^2) + (1-\rho)(1-R_1^2)}{p(1-p)n} \times \frac{(1-p)(c_1 n + c_2) + p(c_1^T n + c_2^T)}{m}. \tag{3}$$

Our framework then identifies the optimal sample allocation under unequal costs by minimizing this error variance of the treatment effect with respect to p (proportion assigned to treatment) and n (individuals/cluster), respectively. The derivations return the following optimal individual sample size

$$n = \frac{\sqrt{(1-\rho)(1-R_1^2)}}{\sqrt{\rho(1-R_2^2)}} \sqrt{\frac{(1-p)c_2 + pc_2^T}{(1-p)c_1 + pc_1^T}}. \tag{4}$$

That is, the optimal number of individuals per cluster is a function of conditional intraclass correlation coefficient, the proportion of groups assigned to treatment, and the costs of sampling (treatment and control) clusters relative to sampling (treatment and control) individuals. More generally, this expression extends the results of conventional equal cost optimal design frameworks in that it appropriately takes into account the differential sampling costs across treatment conditions and levels of the hierarchy.

The consideration of differential costs also introduces the possibility that the unbalanced assignment of clusters to treatment conditions can produce more power. Derivations similar to those above regarding the optimal individual-level sample also suggest that the optimal proportion of clusters to be assigned to the treatment condition is

$$p = \frac{\sqrt{(c_1 n + c_2)\,/(c_1^T n + c_2^T)}}{1 + \sqrt{(c_1 n + c_2)\,/(c_1^T n + c_2^T)}}, \tag{5}$$

That is, given different sampling costs for the treatment and control conditions, the most efficient assignment of clusters to conditions may not be a balanced design but rather one that directly considers the relative cost ratios. The optimal p is driven by the sampling cost ratio between treatment conditions (i.e., $(c_1 n + c_2)\ /(c_1^T n + c_2^T)$), the more expensive it is to sample a cluster and/or its individuals in the treatment condition is, the smaller the optimal p.

Comparison with the Previous Framework

Due to space limitations, we used an example to illustrate that the proposed framework is more efficient (produces more power or requires a smaller budget) than the previous framework (Figures 1-3). Given a set of design parameters (see detail in Figure 1), the proposed framework results in optimal $n = 22$ and optimal $p = .22$, and the previous framework has optimal $n = 32$ and fixed $p = .5$. The individual-level sample size in both frameworks has been in their most efficient allocation (Figure 1). However, the fixed $p = .5$ in the previous framework departs from the most efficient allocation (Figure 2). As a result, the proposed framework produces more power or requires a smaller budget than the previous framework (Figure 3).

Significance

The results presented in this study expand prior considerations regarding how to optimize the design of studies and provide a more flexible set of tools to improve the efficiency and/or rigorousness of these designs. To facilitate end-user calculations, the solutions have been implemented in the free software R package *odr* (Shen & Kelcey, 2018).

References

Glynn, T. J., Shopland, D. R., Manley, M., Lynn, W. R., Freedman, L. S., Green, S. B., ... & Chapelsky, D. A. (1995). Community intervention trial for smoking cessation (COMMIT): I. Cohort results from a four-year community intervention. *American Journal of Public Health, 85*(2), 183-192.

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., ... & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal, 48*(3), 647-717.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87.

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis, 37*(3), 314-332.

Jayanthi, M., Gersten, R., Taylor, M. J., Smolkowski, K., & Dimino, J. (2017). Impact of the developing mathematical ideas professional development program on grade 4 students' and teachers' understanding of fractions (REL 2017–256). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs

Konstantopoulos, S. (2009). Incorporating cost in power analysis for three-level cluster-randomized designs. *Evaluation Review, 33*(4), 335-357.

Konstantopoulos, S. (2011). Optimal sampling of units in three-level cluster randomized designs: An ANCOVA framework. *Educational and Psychological Measurement*, *71*(5), 798-813.

Korendijk, E. J., Moerbeek, M., & Maas, C. J. (2010). The robustness of designs for trials with nested data against incorrect initial intracluster correlation coefficient estimates. *Journal of Educational and Behavioral Statistics, 35*(5), 566-585.

Liu, X. (2003). Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics, 28*(3), 231-248.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2): 113-127.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousands Oaks, CA: Sage.

Springer, M. G., Ballou, D., Hamilton, L., Le, V. N., Lockwood, J. R., McCaffrey, D. F., ... & Stecher, B. M. (2011). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching (POINT).* Society for Research on Educational Effectiveness. Available at https://files.eric.ed.gov/fulltext/ED518378.pdf.

Shen, Z., & Kelcey, B. (2018). *odr: Optimal design and statistical power of multilevel randomized trials* (Version 1.0.0) [Software]. https://cran.r-project.org/web/packages/odr.

Appendix: Figures

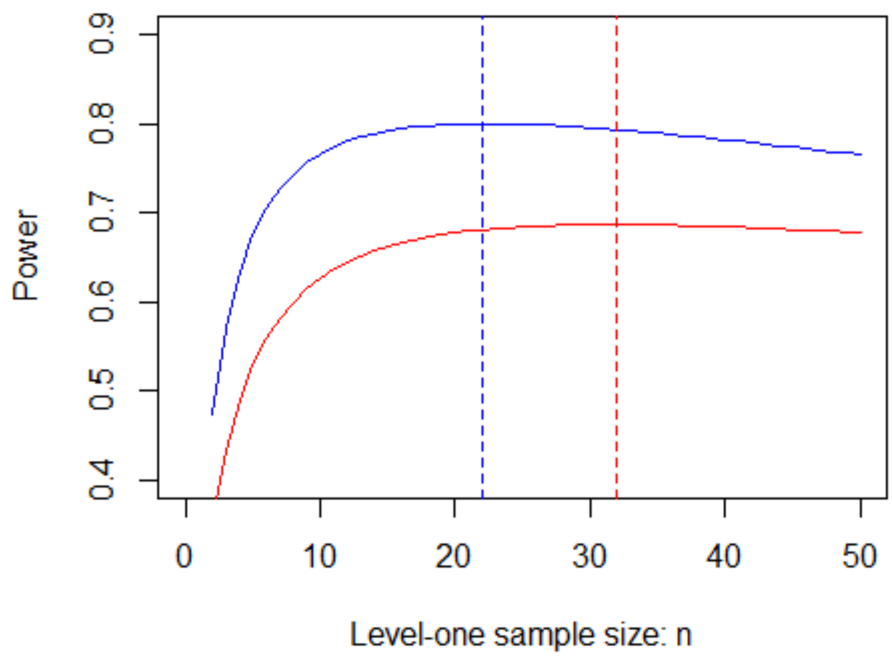

Figure 1. Statistical power of different frameworks under a fixed budget and the effects on power when sampling departs from the optimal n. Red color = the Raudenbush framework with optimal $n = 32$ and **fixed** $p = .5$, blue color = the proposed framework with optimal $n = 22$ and optimal $p = .22$.

Note. Parameter values are $c_1 = 10$, $c_2 = 200$, $c_1^T = 10$, $c_2^T = 5{,}000$, $\rho = .2$ (intraclass correlation coefficient), $R_1^2 = R_2^2 = .5$, $q = 1$ (# of covariates at the level two), $m = 202{,}361$.

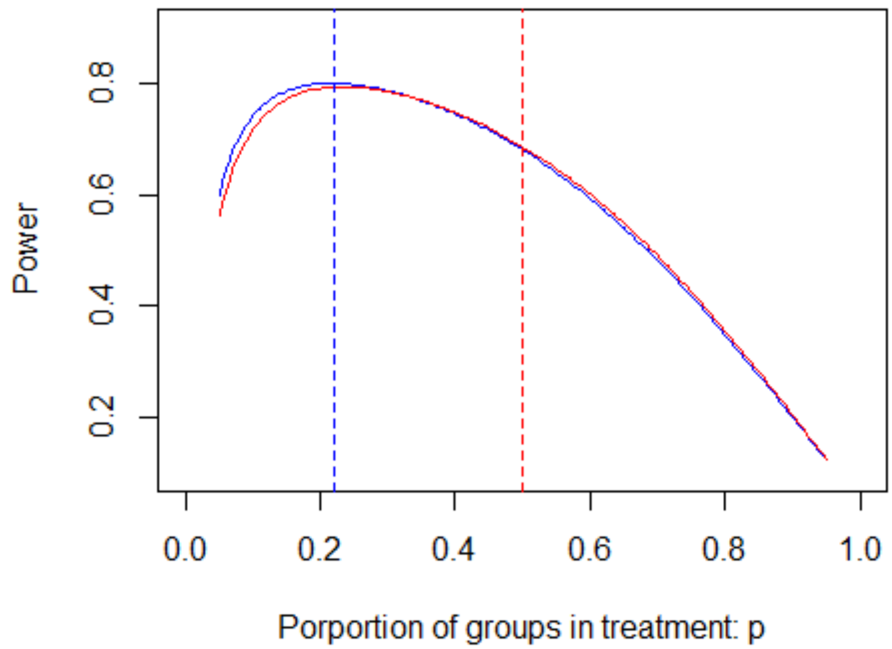

Figure 2. Statistical power of different frameworks under a fixed budget and the effects on power when sampling departs from the optimal/fixed p. Red color = the Raudenbush framework with optimal $n = 32$ and **fixed** $p = .5$, blue color = the proposed framework with optimal $n = 22$ and optimal $p = .22$.

Note. Parameter values are $c_1 = 10$, $c_2 = 200$, $c_1^T = 10$, $c_2^T = 5{,}000$, $\rho = .2$ (intraclass correlation coefficient), $R_1^2 = R_2^2 = .5$, $q = 1$ (# of covariates at the level two), $m = 202{,}361$.

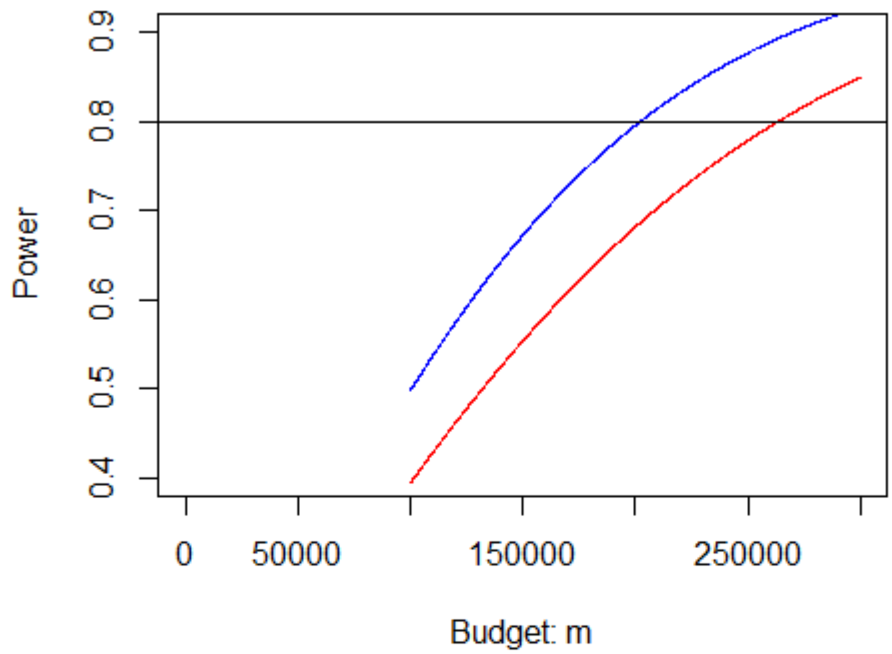

Figure 3. Required budget to achieve a same level of statistical power under different frameworks. Red color = the Raudenbush framework with optimal $n = 32$ and fixed $p = .5$, blue color = the proposed framework with optimal $n = 22$ and optimal $p = .22$.

Note. Parameter values are $c_1 = 10$, $c_2 = 200$, $c_1^T = 10$, $c_2^T = 5{,}000$, $\rho = .2$ (intraclass correlation coefficient), $R_1^2 = R_2^2 = .5$, $q = 1$ (# of covariates at the level two).