Abstract Title Page
Not included in page count.

Title: Optimal Design for Regression Discontinuity Studies with Clustering

Authors and Affiliations: Christopher Rhoads, University of Connecticut.
Charles Dye, University of Connecticut.

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Recent years have seen an increased interest in quantitative educational research studies that use random assignment (RA) to evaluate the causal impacts of educational interventions (Angrist, 2004). The multi-level structure of the public education system in the United States often leads to experimental designs where naturally occurring clusters (eg. schools) are utilized to recruit participants (eg. students) into a study and/or as units that are randomized into one of two experimental conditions.

When multiple clusters are recruited to participate in a research study but individuals within clusters are randomly assigned to experimental conditions the resulting experimental design is often referred to as a *multi-site* design. When clusters are utilized as the unit of randomization the resulting experimental design is generally called a *cluster randomized* or *hierarchical* design. While designs that use some form of random assignment are generally preferred for making causal inferences about treatment effects, RA designs are not always practical or feasible (Shadish, Cook and Campbell, 2002). The next best alternative to a random assignment design is often the regression discontinuity (RD) design. Studies using the regression discontinuity design can be constructed when assignment to treatment occurs as a function of a fully observed *score variable* which has a continuous distribution. Units with scores below a preset cutoff value are assigned to the treatment group, and units with scores above the cutoff value are assigned to the comparison group, or vice versa.

The same clustering issues that arise with random assignment studies also impact regression discontinuity studies. The score variable may be measured at any of many different levels in the educational hierarchy while outcome measurements are obtained at a lower level in the hierarchy. These sorts of situations lead to *hierarchical* regression discontinuity (RD) designs. Alternatively, the units chosen to participate in the study may be clustered in certain sites, even though the score variable is measured within sites. The resulting study is a *multi-site* regression discontinuity design.

An important component of planning experiments with clustering (whether they be multi-site or cluster randomized designs) is to determine the optimal within and between cluster sample sizes subject to a cost constraint. The optimal sample size at each level of the design has been reported for two level cluster randomized designs (Raudenbush, 1997), for two level multi-site randomized designs (Raudenbush and Liu, 2000) and for three level cluster randomized designs (Konstantopoulos, 2009). However, no solution has been provided for the problem of optimal design in regression discontinuity studies.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The current paper builds on previous work by Schochet (2008, 2009) in order to provide formulas for the optimal sample size at each level of a two-level regression discontinuity design.

The paper also provides optimal sample size formulas for unbalanced random assignment designs (such formulas have not previously appeared in the literature). Optimal sample sizes for RD designs are compared to the optimal samples sizes for the corresponding RA designs.

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

Recent work has indicated a resurgence of interest in using regression discontinuity designs for estimating the causal impacts of educational interventions (Cook, 2008; Imbens and Lemieux, 2008). Previous work has provided formulas for computing the variance of impact estimates based on regression discontinuity designs (Schochet, 2009). There has also been work identifying for researchers the optimal sample sizes to use at different levels of the hierarchy in random assignment designs (Raudenbush, 1997; Raudenbush and Liu, 2000; Konstantopoulos, 2009). However, there has not yet been published work illustrating optimal sample sizes for regression discontinuity designs with a hierarchical structure.

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

We assume that the study in question contains *n* clusters, each consisting of *m* individuals who participate in the study. Let Y_{ij} represent the outcome measured on the *j*th individual in the *ith* cluster. In order to ensure comparability of results from RA and RD designs, we assume that the score variable is also measured and included in statistical models for RA designs.

We assume that the researcher has a fixed budget constraint labeled *T*. The marginal cost of including an additional cluster in the study is C_2, and the marginal cost of including an additional individual from a cluster that has already been sampled is C_1. We assume the following simple linear cost function defines the cost constraint:

$$\mathrm{T} = \mathrm{C}_2 n + \mathrm{C}_1 mn \,. \qquad (1)$$

Hierarchical random assignment and regression discontinuity designs.
The standard linear mixed model used to model hierarchical designs can be written as follows:

$$Y_{ij} = \alpha_0 + \alpha_1 T_i + \alpha_2 (Score_{ij} - Score_{i\bullet}) + \alpha_3 Score_{i\bullet} + \tau_i + \varepsilon_{ij} \,. \qquad (2)$$

T_i is a treatment indicator variable coded -0.5 for control cases and 0.5 for treatment cases, and we assume that a proportion, *P*, of the available clusters is randomized to treatment. $Score_{ij}$ is the measured value of the score variable for individual *j* in cluster *i*. $Score_{i\bullet}$ is the average value of the score variable in cluster *i*. The τ_i random variables are random cluster effects and have variance $\sigma^2_{A,\tau}$ and the ε_{ij} random variables are random individual level residuals having variance σ^2_A. The "A" subscripts are included to emphasize that the variance components are "adjusted" in the sense that the variances are computed after conditioning on the score variable. The (adjusted) intracluster correlation coefficient (ICC), ρ_A, is defined as

$$\rho_A = \frac{\sigma^2_{A,\tau}}{\sigma^2_{A,T}} \,. \qquad (3)$$

Here $\sigma^2_{A,T} = \sigma^2_{A,\tau} + \sigma^2_A$ is the total (adjusted) variance in the outcome variable of interest.

Multi-site random assignment and regression discontinuity designs

The regression form of the standard linear mixed model used to model multi-site random assignment designs can be written as follows:

$$Y_{ij} = \alpha_0 + \alpha_1 T_{ij} + \alpha_2 (Score_{ij} - Score_{i\bullet}) + \alpha_3 Score_{i\bullet} + \beta_i + (\alpha\beta)_i T_{ij} + \varepsilon_{ij} \,. \qquad (4)$$

In equation (4) the β_i random variables represent variability in the level of the outcome across clusters and have variance $\sigma^2_{A,\beta}$. The $\alpha\beta_i$ random variables represent variability in the treatment effects across clusters and have variance $\sigma^2_{A,\alpha\beta}$. The "A" subscripts are again included to indicate that the variances are conditional on the score variable. Other quantities are defined as they were in equation (2) with the exception that we now assume a proportion, P, of units <u>within</u> each cluster are randomly assigned to treatment. In order to ensure that the total variance of the outcome variable is the same in the multi-site and cluster randomization formulations of the model it is necessary to make the following equivalence between the variance component parameters

$$\sigma^2_{A,\tau} = \sigma^2_{A,\beta} + \frac{1}{4}\sigma^2_{A,\alpha\beta} \,. \qquad (5)$$

It may be useful to think about the variance in treatment effects relative to the total variance across clusters. As such, we define the parameter ω_A as follows:

$$\omega_A = \frac{\sigma^2_{A,\alpha\beta}/2}{\sigma^2_{A,\tau}} \,. \qquad (6)$$

Usefulness / Applicability of Method:

Demonstration of the usefulness of the proposed methods using hypothetical or real data.

The usefulness of the results is evident from the description in the Findings/Results section.

Findings / Results:

Description of the main findings with specific details.

We use the subscripts "H", "MS", "RA" and "RD" to represent "Hierarchical", "Multi-site", "Random assignment" and "Regression discontinuity", respectively. So, for instance, a result for a multi-site regression discontinuity design would be subscripted "MS, RD". With this notational convention we obtain the following results.

The variance of the treatment effect in a hierarchical RA design is given by the following equation (Bloom, 2005):

$$Var_{H,RA}(\alpha_1) = \sigma^2_{A,T} \frac{1}{P(1-P)} \frac{1}{mn} (1 + (m-1)\rho_A) \,. \qquad (7)$$

When a hierarchical RD design is considered, the equation is modified to read (Schochet, 2008):

$$Var_{H,RD}(\alpha_1) = \sigma^2_{A,T} \frac{1}{(1-\rho^2_{TS})} \frac{1}{P(1-P)} \frac{1}{mn} (1 + (m-1)\rho_A) \,. \qquad (8)$$

The notation " ρ_{TS}^2 " represents the correlation between the score variable and the treatment indicator variable.

Using the method of Lagrange multipliers to optimize equation (7) subject to the constraint given in (2) yields an optimal within-cluster sample size of:

$$m_{opt,H,RA} = \sqrt{\frac{1-\rho_A}{\rho_A}}\sqrt{\frac{C_2}{C_1}}\,. \tag{9}$$

The constrained optimization of equation (8) yields the same result as the constrained optimization of equation (7). This leads us to our first major finding: the optimal within cluster sample size for hierarchical RD designs is that same as the optimal within cluster sample size for hierarchical RA designs. Furthermore, the optimal within cluster sample size does not depend on the proportion of clusters randomized to treatment.

The variance of the treatment effect estimator in multisite RA designs is given by the following equation (Hedges and Rhoads, 2009):

$$Var_{MS,RA}(\alpha_1) = \frac{\sigma^2_{A,\alpha\beta}}{n} + \frac{1}{P(1-P)}\frac{\sigma^2_A}{mn}\,. \tag{10}$$

It is straightforward to derive the corresponding results for multi-site RD designs:

$$Var_{MS,RD}(\alpha_1) = \frac{\sigma^2_{A,\alpha\beta}}{n} + \frac{1}{P(1-P)}\frac{1}{(1-\rho_{TS}^2)}\frac{\sigma^2_A}{mn}\,. \tag{11}$$

Constrained optimization of (10) yields:

$$m_{opt,MS,RA} = \sqrt{\frac{1}{2P(1-P)}}\sqrt{\frac{1-\rho_A}{\omega_A\rho_A}}\sqrt{\frac{C_2}{C_1}}\,, \tag{12}$$

And constrained optimization of (11) yields:

$$m_{opt,MS,RD=} = \sqrt{\frac{1}{2P(1-P)(1-\rho_{TS}^2)}}\sqrt{\frac{1-\rho_A}{\omega_A\rho_A}}\sqrt{\frac{C_2}{C_1}}\,. \tag{13}$$

In contrast to hierarchical designs, in multi-site designs the optimal within cluster sample size will depend on both the proportion of units randomized to treatment and on the extent of the correlation between the treatment indicator variable and the score variable in RD designs. Designs that are highly unbalanced will result in larger within cluster sample sizes, as will designs where the value of ρ_{TS}^2 is large. A sampling of the optimal within cluster sample sizes obtained under the various design scenarios considered is presented in Table 1 in the appendix. Results are presented for various cost ratios (values of $\frac{C_2}{C_1}$), proportions in treatment (*P*), values of the conditional ICC (ρ_A) and values of the treatment effect heterogeneity parameter (ω_A).

Conclusions:

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

This study has shown that the optimal within cluster sample size for a two level hierarchical design is the same regardless of whether the study is a regression discontinuity study or a random assignment study. The optimal sample size in hierarchical designs is also insensitive to differences between the number of clusters in the two conditions being compared.

The story is different when multi-site designs are considered. Regression discontinuity studies should be structured to have more units within each site than random assignment designs. Additionally, if the design is unbalanced (more units in one condition than the other) the within cluster sample size should be increased.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Bloom, H.S. (2005). Randomizing groups to evaluate place-based programs. In H.S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 115-172). New York: Russell Sage Foundation.

Cook, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics,* 142, 636–54.

Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615-635.

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

Raudenbush, S.W. (1997). Statistical analysis and optimal design for cluster randomization trials. *Psychological Methods,* 2, 173-85.

Raudenbush, S.W. and Liu, X. (2000). Statistical analysis and optimal design for multisite randomized trials. *Psychological Methods,* 5, 199-213.

Schochet, Peter Z. (2008). *Technical Methods Report*: *Statistical Power for Regression Discontinuity Designs in Education Evaluations* (NCEE 2008-4026). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schochet, P.Z. (2009). Statistical power for regression discontinuity designs in education evaluations. *Journal of Educational and Behavioral Statistics*, 34(2), 238-266.

Appendix B. Tables and Figures

Table 1: *Optimal within cluster sample sizes.*

Cost ratio	ICC	*P*	ω	Hier opt. *m*	MS opt. *m* (RA)	MS opt. *m* (RD)
2	0.05	0.5	0.05	6	39	65
2	0.05	0.5	0.5	6	12	21
2	0.05	0.3 (or 0.7)	0.05	6	43	66
2	0.05	0.3 (or 0.7)	0.5	6	14	21
2	0.05	0.1 (or 0.9)	0.05	6	65	80
2	0.05	0.1 (or 0.9)	0.5	6	21	25
2	0.15	0.5	0.05	3	21	36
2	0.15	0.5	0.5	3	7	11
2	0.15	0.3 (or 0.7)	0.05	3	23	36
2	0.15	0.3 (or 0.7)	0.5	3	7	11
2	0.15	0.1 (or 0.9)	0.05	3	36	44
2	0.15	0.1 (or 0.9)	0.5	3	11	14
2	0.25	0.5	0.05	2	16	26
2	0.25	0.5	0.5	2	5	8
2	0.25	0.3 (or 0.7)	0.05	2	17	26
2	0.25	0.3 (or 0.7)	0.5	2	5	8
2	0.25	0.1 (or 0.9)	0.05	2	26	32
2	0.25	0.1 (or 0.9)	0.5	2	8	10
5	0.05	0.5	0.05	10	62	103
5	0.05	0.5	0.5	10	20	33
5	0.05	0.3 (or 0.7)	0.05	10	67	104
5	0.05	0.3 (or 0.7)	0.5	10	21	33
5	0.05	0.1 (or 0.9)	0.05	10	103	126
5	0.05	0.1 (or 0.9)	0.5	10	33	40
5	0.15	0.5	0.05	5	34	56
5	0.15	0.5	0.5	5	11	18
5	0.15	0.3 (or 0.7)	0.05	5	37	57
5	0.15	0.3 (or 0.7)	0.5	5	12	18
5	0.15	0.1 (or 0.9)	0.05	5	56	69
5	0.15	0.1 (or 0.9)	0.5	5	18	22
5	0.25	0.5	0.05	4	25	41
5	0.25	0.5	0.5	4	8	13
5	0.25	0.3 (or 0.7)	0.05	4	27	41
5	0.25	0.3 (or 0.7)	0.5	4	9	13
5	0.25	0.1 (or 0.9)	0.05	4	41	50
5	0.25	0.1 (or 0.9)	0.5	4	13	16
15	0.05	0.5	0.05	17	107	178
15	0.05	0.5	0.5	17	34	56
15	0.05	0.3 (or 0.7)	0.05	17	117	179
15	0.05	0.3 (or 0.7)	0.5	17	37	57
15	0.05	0.1 (or 0.9)	0.05	17	178	218
15	0.05	0.1 (or 0.9)	0.5	17	56	69
15	0.15	0.5	0.05	9	58	97
15	0.15	0.5	0.5	9	18	31
15	0.15	0.3 (or 0.7)	0.05	9	64	98
15	0.15	0.3 (or 0.7)	0.5	9	20	31
15	0.15	0.1 (or 0.9)	0.05	9	97	119
15	0.15	0.1 (or 0.9)	0.5	9	31	38
15	0.25	0.5	0.05	7	42	71
15	0.25	0.5	0.5	7	13	22
15	0.25	0.3 (or 0.7)	0.05	7	46	71
15	0.25	0.3 (or 0.7)	0.5	7	15	23
15	0.25	0.1 (or 0.9)	0.05	7	71	87
15	0.25	0.1 (or 0.9)	0.5	7	22	27

Note: Table assumes a normally distributed score variable