SREE

2040 Sheridan Rd
Evanston, IL 60208

Society for Research on Educational Effectiveness
Advancing Education Research

www.sree.org
847.467.4001

2010 SREE Conference Abstract Template

Thank you for your interest in the Society for Research on Educational Effectiveness 2010 Annual Conference. Conference abstracts must be submitted using this template document. The template is based on the recommendations offered by Mosteller, Nave, and Miech (2004, p. 33)[1] for structured abstracts. Abstracts should follow APA Style, as specified in the Sixth Edition of the Publication Manual of the American Psychological Association.

Abstract Preparation and Submission Procedures

Save this document to your computer. Fill in each of the sections below with the relevant information about your proposed poster, paper, or symposium. Make sure to save the document again when completed. When ready, submit your abstract at http://www.sree.org/conferences/2010/submissions/

The template consists of the following sections: title page, abstract body, and appendices (references and tables and figures). Figures and tables included as part of submission should be referred to parenthetically—"(please insert figure 1 here)." The body section of your abstract should be no longer than 5 pages (single spaced, using the Times New Roman 12-point font that has been set for this document). The title page and appendices do not count toward this 5-page limit.

Insert references in appendix A of this document. Insert tables and graphics in appendix B. Do not insert them into the body of the abstract.

**For questions, or for help with abstract preparation or submission,
contact us at inquiries@sree.org, or 847-467-4001**

1 Mosteller, F., Nave, B., & Miech, E. (2004). Why we need a structured abstract in education research. *Educational Researcher, 33(1),* 29–34.

Abstract Title Page

Not included in page count.

Title: **The Trade-off of Bias versus Power in Regression Discontinuity Design Estimators**

Author(s): **Rebecca Maynard**
Nirav Mehta
Nianbo Dong

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Regression Discontinuity (RD) design has a long history in program evaluation (see Trochim 1984 and Cook 2008), and remains a popular method for evaluating programs using observational datasets in which the program is assigned with different probabilities above and below a known cut-off. There are two typical setups: Sharp and Fuzzy. In Sharp designs, the probability an individual is assigned to the program below the cut-off is 0 (or 1), and the probability of being assigned to the program above the cutoff is the complementary probability of 1(or 0). In Fuzzy designs, the probability of receiving the program must be discontinuous, or jump, at the cutoff. We focus on Sharp designs, which we hereafter refer to as RD design. To fix ideas, take the following example of a Sharp design: All children with a grade point average below a known threshold are assigned to summer school, while those with grade point averages above the threshold are not assigned to summer school. We can then compare subsequent outcomes for students *just below* and *just above* the threshold (i.e. those assigned to summer school and not assigned to summer school) to evaluate the effect of the program (in this case, the program is mandatory summer school). Students far away from the cut-off are potentially different along unobserved dimensions in addition to their participation in the program, thereby confounding estimates of the treatment effect based on a simple regression analysis using the full sample.

A key feature of RD design is that it is that the likelihood that estimates of the treatment effect will be unbiased decreases as the distance from the cut-off increases. This means that RD design may deliver an unbiased estimate of the Local Average Treatment Effect at the expense of greatly reducing power, since only individuals just above and just below the cutoff (i.e. those "local" to the cut-off) can be used in calculating the treatment effect. One response to this power reduction has been to include in the analysis individuals further from the cut-off, which, in the limit where one includes the full sample, amounts to a linear regression with the program indicator included as a regressor. This approach has "more power," but increases the potential bias. Put another way, it effectively reduces the design to a more typical regression-control design.

There are several methodological papers on RD design, but most focus on analyzing the bias induced by mis-specification errors, for example, adding a linear treatment-effect term in a RD design model, thereby using individuals far from the cut-off to evaluate the effect at the cut-off (Hahn, Todd, and Van der Klaauw 2001). The problem is that the very reasons for using RD designs mean that it is not possible to reliably test all of the underlying assumptions that support treating resulting impact estimates as unbiased. This methodological paper can help inform policymakers by clarifying which RD design estimators are appropriate in a given context.

In this paper, we investigate the trade-off between increasing power and introducing bias by estimating treatment effects using common RD estimators under different scenarios of unobservable heterogeneity and treatment effect types.

Purpose / objective / research question / focus of study:
Description of what the research focused on and why.

We provide empirical illustrations of the performance of RD designs applied in prototypical manners with a nationally-representative data set, the ECLS-K. Our primary goal is to estimate bias in impact estimates and standard errors under various known conditions with respect to the relationships among the decision rule for assignment to treatment condition, the size and variability in the true impact, and unobserved heterogeneity of the sample.

We generate data using the ECLS-K to simulate how different combinations of treatment effects, unobservable heterogeneity, and selection into the treatment group affect different types of RD estimators in both terms of bias and power. In particular, we compare the bias and power of RD estimators that are inherently local (caliper and non-parametric local linear regression methods) with those that use a observations further from the cutoff – in potential violation of the assumptions of RD design. We also show how RD design can create a biased estimate of the treatment effect if applied to individuals not near the cut-off – that is how it is a Local Average Treatment Effect instead of an Average Treatment Effect, which is what the experimental estimator recovers. These two estimators of the treatment effect would differ if the size of the treatment effect was a function of the variable inducing the cut-off. For example, a reading program may be assigned based on a reading test score cutoff, which occurs at the median reading test score. If the effectiveness of the program depends on the initial score, however, then applying the RD estimator of the treatment effect to individuals not around the cutoff can lead to an incorrect assessment of the treatment effect. We also demonstrate how the RD design can be biased by the inclusion of individuals far from the cut-off in the case where there are unobservable differences that affect both the treatment effect and selection into the program. The Sharp RD estimator avoids bias in this case, but does so with a reduction of power, so we further demonstrate this trade-off under scenarios involving different data generating processes.

Setting:
Description of where the research took place.

Observational dataset (ECLS-K).

Population / Participants / Subjects:
Description of participants in the study: who (or what) how many, key features (or characteristics).

We use simulated data based on the ECLS-K.

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

We simulate an intervention in this methodological paper.

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

This is a methodological paper about Regression Discontinuity design, which is often classified as a "Quasi-Experimental" design.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

Observational dataset (ECLS-K).

Findings / Results:
Description of main findings with specific details.

We demonstrate the trade-off between increasing potential bias and reducing power under commonly used Regression Discontinuity design methods. This is further discussed in the next section.

Conclusions:
Description of conclusions and recommendations based on findings and overall study.

Depending on the underlying nature of unobserved heterogeneity and assumptions about how the treatment effect depends on individual characteristics, the estimates derived from RD design can vary significantly, depending on the method used. We show that local methods, i.e. those using the portion of the sample only near the cut-off, provide the best estimates of the Local Average Treatment Effect in that they are robust to various assumptions about unobservable heterogeneity and how the treatment affects the outcomes, but they also greatly diminish the power of the estimator. We then perform a power analysis of different RD design estimators and compare these with bias in estimating the treatment effect under various scenarios.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Bloom, H. and Kemple, J. and Gamse, B. and Jacob, R. (2005) *Using regression discontinuity analysis to measure the impacts of Reading First.* Paper presented at the annual conference of the American Educational Research Association, Montreal Canada, 2005.

Cook, T.D. (2008) "*Waiting for Life to Arrive": A history of the regression-discontinuity design in Psychology*, *Statistics and Economics.* Journal of Econometrics, 142:2, 636—654.

Hahn, J. and Todd, P. and Van der Klaauw, W. (2001) *Identification and estimation of treatment effects with a regression-discontinuity design.* Econometrica, 69, 201-209.

Trochim, W.M.K. (1984). *Research design for program evaluation: The regression-discontinuity approach.* Beverly Hills, CA: Sage Publications, Inc.

Appendix B. Tables and Figures

Not included in page count.