Design Considerations in Three-level Regression Discontinuity Studies

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
at the University of Missouri - Columbia

by

Metin Bulus

Dr. Nianbo Dong, Dissertation Co-supervisor
Dr. Ze Wang, Dissertation Co-supervisor

December 2017
Design Considerations in Three-level Regression Discontinuity Studies

presented by Metin Bulus,

Dr. Nianbo Dong

Dr. Ze Wang

Dr. Francis L. Huang

Dr. Wolfgang Wiedermann

Dr. Phillip K. Wood
Acknowledgements

During the little over five years of my stay at the University of Missouri, I have met many great faculty, staff, and student colleagues who have supported me on this journey to become a scholar. Among them is my co-advisor, supervisor, colleague and friend, Dr. N. Dong to whom I express my deepest gratitude. From the conception of this dissertation to its completion, Dr. N. Dong’s guidance has been vital, and the impact of his mentorship has been immeasurable.

I would also like to express my gratitude to my other co-adviser Dr. Z. Wang, who also believed in me and guided me along my journey as a scholar. Special thanks to my committee members Dr. P. K. Wood, Dr. F. L. Huang, and Dr. W. Wiedermann who gave me crucial feedback at various stages of the dissertation.

My colleague Mr. P. Shao’s insight and help in deriving the formulas has been an important step to achieve this goal. Many thanks to the ESCP faculty, staff and colleagues who were always there at times in need and made me feel at home.

Additionally, there are many teachers throughout my life, to whom I owe who I am today. From my elementary school to university education, my teachers have never reserved their advice, wisdom and support. I would like to thank Dr. B. Cavas, Dr. K. Demir, Dr. A. B. Koc, Dr. S. Erol, Mr. I. Gunes, Mr. L. Simsek, and Mr. U. Uyar for being my teachers for life.

I would also like to thank my government; especially the Turkish Ministry of Education for providing me an opportunity to study abroad, which otherwise would not
have been possible. Special thanks to the Turkish Education Attaché in New York whose support has been immeasurable.

Finally, yet importantly, my wife and my son has been the constant source of love, support, and inspiration. Thank you for being with me through the ups and downs of this journey. This work is indeed dedicated to both of you.
# Table of Contents

Acknowledgements........................................................................................................................................ ii

List of Tables ................................................................................................................................................... ix

List of Figures ................................................................................................................................................ x

Abstract......................................................................................................................................................... xi

1 Introduction .................................................................................................................................................. 11

1.1 Background ............................................................................................................................................. 11

1.1.1 RD Designs in Education.................................................................................................................. 13

1.1.2 Statistical Power in RE Designs in Education ................................................................................. 14

1.1.3 Statistical Power in RD Designs in Education............................................................................... 15

1.2 Problem Statement ................................................................................................................................. 19

1.3 Significance and Contribution of the Current Study ........................................................................... 21

2 Literature Review ..................................................................................................................................... 24

2.1 Theoretical Underpinnings.................................................................................................................... 24

2.2 Type I and Type II Error Rates ............................................................................................................. 33

2.3 MDE(S), MRSS and Statistical Power..................................................................................................... 35

2.4 Relative Efficiency and Statistical Power in RD Design as opposed to RE ..................................... 39

2.5 Statistical Power Analysis for RD Design with Clustering................................................................. 42

2.5.1 Two-level Blocked Individual RD Designs with Fixed Effects (RD2ff1) ..................................... 44
2.5.2 Two-level Blocked Individual RD Designs with Random Effects (RD2rr1) 46
2.5.3 Two-level Cluster RD Designs with Random Effects (RD2r2) .......... 48
2.5.4 Three-level Cluster RD Designs with Random Effects (RD3r3) ....... 50
2.5.5 Three-level Blocked Cluster RD Designs with Fixed Effects (RD3ff2) .... 52
2.5.6 Three-level Blocked Cluster RD Designs with Random Effects (RD3rr2) 54

2.6 Consequences of Ignoring a Level of Nesting Structure ....................... 57

2.6.1 Effects of Ignoring a Level of Nesting on Variance Components and their
Standard Errors ......................................................................................... 58

2.6.2 Effects of Ignoring a Level of Nesting on Fixed Effect Estimates and their
Standard Errors ......................................................................................... 60

2.6.3 Effects of Ignoring a Level of Nesting in Unbalanced Designs compared to
Balanced Designs ....................................................................................... 62

2.7 Evidence from Empirical Studies that Ignore a Level of Nesting in RD Design 63

2.8 Constrained Optimal Sample Allocation .................................................. 65

2.9 Gap in the Literature and Significance of the Current Study ..................... 67

2.9.1 Gap in Statistical Power Analysis for Clustered RD Studies ............... 68

2.9.2 Gap in Consequences of Ignoring Intermediate- or Top-level in Clustered
RD Studies ..................................................................................................... 69

2.9.3 Gap in Optimal Design .......................................................................... 70

3 Method ........................................................................................................ 73
3.1 Derivation of Sampling Variance ................................................................. 73

3.2 Models and Monte Carlo Simulations .......................................................... 78

3.2.1 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
and Level 3 Intercepts and Treatment Effect are Random (BIRD3rr2rr1) .......... 78

3.2.2 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
and Level 3 Intercepts are Random but Treatment Effect is Constant (BIRD3rc2rc1) 85

3.2.3 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts Mean Effects are Random but Treatment Effect is Random and Level 3
Mean Effects are Fixed but Treatment Effect is Constant (BIRD3fc2rr1) .......... 89

3.2.4 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts are Random but Level 3 Intercepts are Fixed, and Treatment Effect is
Constant across Level 2 and Level 3 (BIRD3fc2rc1) ........................................ 94

3.2.5 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts and Level 3 Intercepts are Fixed and Treatment Effect is Constant
(BIRD3fc2fc1) .................................................................................................... 97

3.3 Consequences of Ignoring Intermediate- or Top-Level ............................... 101

3.4 Constrained Optimal Sample Allocation ..................................................... 103

3.4.1 Framework ............................................................................................. 104

4 Results ............................................................................................................. 110

4.1 Derivation and Validation of Formulas ....................................................... 111
4.1.1 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts and Level 3 Intercepts and Treatment Effect are Random (BIRD3rr2rr1)

111

4.1.2 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts and Level 3 Intercepts are Random (RD3rr2rr-1) but Treatment Effect is
Constant (BIRD3rc2rc1).................................................................................... 124

4.1.3 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts Mean Effects are Random but Treatment Effect is Random and Level 3
Mean Effects are Fixed but Treatment Effect is Constant (BIRD3fc2rr1).......... 128

4.1.4 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts Mean Effects are Random but Treatment Effect Constant and Level 3
Mean Effects are Fixed but Treatment Effect is Constant (BIRD3fc2rc1)......... 133

4.1.5 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2
Intercepts and Level 3 Mean Effects are Fixed and Treatment Effect is Constant
(BIRD3fc2fc1)..................................................................................................... 138

4.2 Omission of Intermediate Level in BIRD3rr2rr1 Model ......................... 142

4.3 Constrained Optimal Sample Allocation................................................. 147

4.3.1 Introduction to PowerUpR R Package.................................................. 147

4.3.2 Example .................................................................................................. 148

5 Conclusion..................................................................................................... 154

References........................................................................................................ 170
Appendix A ......................................................................................................................... 184
Appendix B ......................................................................................................................... 194
Vita ........................................................................................................................................ 206
List of Tables

Table 1.1 Characteristics of Schochet’s (2008b) Six RD Designs and Five Proposed Designs.................................................................................................................................................. 21
Table 2.1 Type I and Type II Error Rates in Hypothesis Testing.......................... 34
Table 2.2 Characteristics of COSA in Existing Literature and this Study............. 72
Table 4.1 Power Analysis for BIRD3rr2rr1 Design .............................................. 115
Table 4.2 Type I Error Analysis for BIRD3rr2rr1 Design................................. 116
Table 4.3 Power Analysis for the Misspecified BIRD2rrr1 Model (Level 2 Ignored) .... 143
Table 4.4 Power Analysis for the Misspecified BIRD2rrr1 Model (Level 3 Ignored) .... 144
Table 4.5 Estimated Variance Parameters from Correctly Specified and Misspecified Unconditional Models.......................................................................................................................... 146
Table 5.1 Statistical Power for BIRD3rr2rr1 Design ........................................... 157
Table 5.2 Minimum Detectable Effect Size for BIRD3rr2rr1 Design................. 159
Table 5.3 Number of Top-level Sample Size for BIRD3rr2rr1 Design............... 161
Table 5.4 Number of Top-level Sample Size for BIRD3rc2rc1 Design............... 163
Table 5.5 Comparison of COSA Solutions in PowerUpR and Optimal Design Plus (OD+) for a Two-level Cluster Randomized Controlled Trial ................................................................. 167
List of Figures

Figure 2.1 Regression discontinuity (binwidth=.05, number of bins=95).......................... 31

Figure 2.2 Multiplier as a function of central and non-central t distributions for one-tailed test................................................................. 37
Abstract

In education, sample characteristics can be complex due to the nested structure of students, teachers, classrooms, schools, and districts. In the past, not many considerations were given to such complex sampling schemes in statistical power analysis. More recently in the past two decades, however, education scholars have developed tools to conduct statistical power analysis in randomized experiments (RE) and regression discontinuity (RD) studies considering complex sampling schemes.

The purpose of this study is threefold: (i) to derive formulas for various three-level RD studies where discontinuity resides at level 1 and to validate formulas using Monte Carlo (MC) simulations, (ii) to explore consequences of ignoring an intermediate- (e.g., classroom/teacher) or top-level (e.g., school/district) when designing such studies, and (iii) to provide a general framework for calculating optimal sample sizes under budget and sample size constraints when treatment and control units are associated with certain costs (equal or unequal).

Derived formulas are consistent with the current literature and uses parameters commonly reported in the education studies. MC simulation results confirm validity of the formulas. On the one hand, ignoring an intermediate-level result in under-powered studies and is not recommended. An intermediate-level may be ignored had the variance of the outcome between level 2 units been small. On the contrary, ignoring top-level result in over-powered studies and is not recommended. In this case, Type I errors are severely inflated, therefore, a researcher is more likely to detect a treatment effect when in reality there is not. Finally, the general framework for constrained optimal sample allocation allows calculation of sample sizes under budget and sample size constraints.
when treatment and control units are associated with certain cost (equal or unequal).

When cost associated with each unit depend on the treatment membership, the proportion of units in treatment condition ($P$) can also be optimized in multilevel RE studies.
1 Introduction

1.1 Background

Randomized experiments (RE) have been considered the gold standard to draw causal inferences with human subjects. In some instances, for efficiency and ethical reasons, RE may not be feasible, in particular, where a treatment is based on the need or merit of subjects. Treatment status considerations based on neediness or meritocracy is common in education, ranging from Head Start program eligibility based on family income level, to scholarship programs where considerations are based on grade point averages. Such high-stake programs require a thorough evaluation of their effectiveness before possible extensions or scale-ups. For example, stake-holders may wish to find out whether there is a meaningful effect attributed to Head Start or scholarship program on student achievement and well-being. In these studies, any degree of randomization may raise ethical issues, furthermore, randomization may not be feasible due to pre-defined target population (low income students and successful students).

When randomization is not possible scholars sought alternative methods to draw causal inferences. Since Thistewhite and Campbell (1960) reanalyzed a group of students who were near winners in a national scholarship competition using regression discontinuity (RD) design and disapproved findings from previous matching studies, RD design has found a wide range of applications from education, psychology to economics field. A few decades after Thistewhite and Campbell introduced RD design, interest in this method faded in education, but began gaining popularity in the beginning of the 21st
century among scholars in education as well as scholars in economics. For a multidisciplinary historical development of RD design see the review by Cook (2008).

The rationale behind the development of RD design and its analogy to RE emerge from assignment mechanisms. In both methods assignment mechanism is fully known in advance. While the assignment mechanism in RE is completely random, in RD design the assignment mechanism is completely controlled where subjects below or above a cutoff score on the assignment variable is granted the treatment. Since the assignment mechanism is fully known in advance, bias that may arise from the selection of subjects can be addressed through statistical modeling. Had the assignment mechanism been uncontrolled, that is when the selection of subjects to treatment groups are partially determined by confounders, some other quasi-experimental methods such as propensity score matching (Rosenbaum & Rubin, 1983) could be utilized. The common goal of these methodologies is to achieve a state of orthogonality between the treatment status and the possible confounders, referred to as un-confoundedness, a key assumption for valid causal inference (Heckman & Hotz, 1989; Rosenbaum & Rubin, 1983). To achieve un-confoundedness, methods other than RE rely on statistical adjustments in the analysis phase as opposed to design phase, and requires stronger modeling assumptions. Although RE and RD design are different with respect to their assignment mechanisms, considering a very narrow range of scores around the cutoff score in RD design resembles an RE due to fluctuations resulting from random measurement error (Boruch, 1975; Campbell & Stanley, 1963; Lee & Lemieux, 2010). The narrower of a range is inspected around the cutoff score, the more random measurement error determines which subject is granted treatment. For a general methodological treatment of the RD method see Bloom (2012),
Cappelleri (1991), and Trochim (1984); for a comprehensive methodological treatment see Imbens and Lemieux (2008).

1.1.1 RD Designs in Education

At the beginning of the 21st century, a plethora of studies in education utilized RD design to evaluate programs. Some well-known studies include, but are not limited to, the impact of response to intervention practices on elementary school reading (Balu et al., 2015), the double-dose algebra program on math achievement (Cotes, Goodman, & Nomi, 2015), the transition to algebra program on students’ mathematics outcomes (Louie, Rhoads, & Mark, 2016), the Head Start program on students’ achievement and health (Ludwig & Miller, 2007), the state voluntary Pre-K program on students’ academic achievement (Lipsey, Farran, Bilbrey, Hofer, & Dong, 2011), the summer school attendance on math and reading (Matsudaira, 2008), the reading recovery program on students’ reading (May, Sirinides, Gray, & Goldsworthy, 2016), and the preschool programs’ on children’s early literacy skills (Wong, Cook, Barnett, & Jung, 2008).

The resurrection of RD design in education warranted setting standards to guide researchers along with RE. In the What Works Clearinghouse (WWC) Handbook (U.S. Department of Education, 2014), RD design is included as one of the credible methods for evaluation after RE and is given two points1 on the internal validity (three for RE, one for quasi-experimental designs2). In addition, to meet the WWC standards without

---

1 Possible points range from zero to three. Higher points indicate higher internal validity.
2 In this study RD studies are categorized under the umbrella of designs where assignment mechanism is fully known. Therefore they can be designed in advance, and a priori power analysis applies. It is not considered to be one of the quasi-experimental design.
reservation, standard errors of the treatment effect estimate in the clustered RD case should be properly addressed by one of the accredited methods such as bootstrapping, hierarchical linear modeling (Raudenbush & Bryk, 2002) or the method proposed by Lee and Card (2008) to address misspecification errors. Clustering is common in education due to the nesting of students in classrooms or teachers, nesting of classrooms or teachers in schools, and nesting of schools in districts. Any incomplete consideration of the data structure at the analysis phase may have adverse consequences on variance components, estimates, and their standard errors (Moerbek, 2004; Opdenakker & Van Damme, 2000; Van den Noortgate, Opdenakker, & Onghena, 2005).

1.1.2 Statistical Power in RE Designs in Education

The standard error of a treatment effect estimate is directly related to statistical power, as under repeated sampling the ratio of the estimate to its standard error follows a non-central $t$ distribution (Cohen, 1988). Had there been a nil treatment effect, with repeated sampling, the ratio of the estimate to its standard error would have followed a central $t$ distribution. Within a frequentist framework, hypothesis tests are conducted via comparing the non-central $t$ distribution to the central $t$ distribution. With repeated sampling, a researcher may wish to set some accuracy level to detect treatment effect if any exists. In this case, the accuracy level is referred to as statistical power. Statistical power is sample and model dependent; therefore, a standard can be set prior to

---

3 Treatment effect is defined as adjusted or unadjusted difference between two groups.
conducting a study by determining the sample size needed to achieve a certain level of precision while presuming some modeling assumptions.

In education, sample characteristics can be complex due to the nesting structure of students, teachers, classrooms, schools, and districts. In the past, not many considerations were given to such complex sampling schemes in statistical power analysis. In the past two decades, however, education scholars have developed tools to conduct statistical power analysis in RE and RD design considering complex sampling schemes (e.g., Bloom, 1995; Bloom, 2006; Bloom, Bos, & Lee, 1999; Bloom, Richburg-Hayes, & Black, 2007; Dong, Kelcey, & Spybrook, 2017; Dong & Maynard, 2013; Hedges & Rhoads, 2010; Kelcey, Dong, Spybrook, & Cox, 2017; Kelcey, Dong, Spybrook, & Shen, 2017; Konstantopoulos, 2009a, 2009b, 2011, 2013a, 2013b; Schochet, 2005, 2008a, 2008b, 2009; Spybrook, Kelcey, & Dong, 2016).

1.1.3 Statistical Power in RD Designs in Education

Although both RE and RD design produce unbiased estimates of the treatment effect, under accurate modeling assumptions RD design requires much larger sample sizes to reach the same level of efficiency as RE due to differences resulting from the assignment mechanism. Without any consideration of complex sampling schemes, Goldberger (1972a, 1972b) found that RD design requires 2.75 times as many participants as an RE in the most extreme case, assuming a normally distributed assignment variable with a cutoff at the mean. On the other hand, considering complex sampling schemes, Schochet (2008b, 2009) showed RD design requires three to four times as many participants as an RE study would require, assuming various distributions
for the assignment variable (normal, uniform, bimodal, and truncated normal). Schochet (2008b) has noted that the difference arises from distributional assumptions rather than complex sampling schemes. The complexity of the sampling schemes can be addressed similarly both in RE and RD studies using either bootstrapping, hierarchical linear modeling (Raudenbush & Bryk, 2002) or Lee and Card’s (2008) method. On the other hand, assignment mechanism in RD studies generate a treatment status strongly correlated with the assignment variable. The degree of this correlation relies on the distribution of the assignment variable which may determine the additional sample size required for RD studies relative to RE studies to obtain the same level of efficiency.

Schochet (2008b) considered six complex sampling schemes for statistical power analysis in RD design (i) two-level designs where subjects are assigned to treatment and control conditions within level 2 units (sites or blocks), (ii) two-level designs where subjects are assigned to treatment and control conditions within sites, (iii) two-level designs where clusters are assigned to treatment and control conditions, (iv) three-level designs where clusters are assigned to treatment and control conditions, (v) three-level designs where level 2 clusters are assigned to treatment and control conditions within level 3 units (sites or blocks), and (vi) three-level designs where level 2 clusters are assigned to treatment and control conditions within level 3 units (sites or blocks). It should be noted that treatment status is determined based on a subject’s location on an assignment variable in comparison to a predetermined cutoff score, unlike RE where the treatment status is determined randomly. For each of the afore-mentioned designs, Schochet (2008b) derived sampling variance of the treatment effect estimate to be used in statistical power analysis. Comparing these sampling variances to RE cases, Schochet
(2008b) found that the inflation in required sample sizes emerge from the relationship between treatment status and assignment variable. The two variables are assumed to have no relationship in RE case, in other words, the assignment variable is just another covariate to be included in the model that bears no relation to either the treatment status or the outcome.

Sampling variance of the treatment effect estimate in simple or clustered RD design differ from simple or clustered RE by an efficiency factor formulated as

\[
D = \frac{1}{1-\rho_{TZ}^2}
\]

(1.1)

where \(\rho_{TZ}^2\) is the squared correlation between the assignment and the treatment variables (Bloom, 2012; Goldberger, 1972a, 1972b, Capilleri 1991, 1994; Schochet 2008b, 2009). Denoting the treatment effect estimate for RD design as \(\delta_{RD}\) and for RE as \(\delta_{RE}\), else being identical, the relationship between the sampling variances for the estimates under these two designs is

\[
\text{var}(\delta_{RD}) = D\text{var}(\delta_{RE})
\]

(1.2)

By having the sampling variance, one can obtain statistical power using a non-central \(t\) distribution, with a non-centrality parameter formulated as (Hedges & Rhoads, 2009)
\[
\lambda = \frac{\delta_{RD}}{\sqrt{\text{var}(\delta_{RD})}}
\] (1.3)

or minimum detectable effects (MDE) formulated as (Bloom, 1995, 2006)

\[
\text{MDE}(\delta_{RD}) = M_v \sqrt{\text{var}(\delta_{RD})}
\] (1.4)

or minimum detectable effect size (MDES as

\[
\text{MDES}(\delta_{RD}) = M_v \sqrt{\text{var}(\delta_{RD})/\sigma_F^2}
\] (1.5)

where \(\sigma_F^2\) is the variance of the outcome and \(M_v = t_\alpha + t_\beta\) for one tailed test and \(M_v = t_{\alpha/2} + t_\beta\) for a two-tailed test. Critical \(t\) values are calculated based on a type I error rate (\(\alpha\)) and a type II error rate (\(\beta\)) (Bloom, 1995, 2006; Dong & Maynard, 2013). Type I and Type II error rates will be discussed in following sections. Accordingly, sample sizes can be obtained by inverting any of these functions. The framework remains the same, however, with complex sampling schemes \(D\) factors in differently depending on the level of discontinuity.
1.2 Problem Statement

Schochet (2008b) considered individual level assignment with two-level designs, however, three-level designs with the assignment variable at the individual level is prevalent (e.g., Gamse, Jacob, Horst, Boulay, & Unlu, 2008; Henry, Fortner, & Thompson, 2010; Hustedt, Jung, Barnett, & Williams, 2015). Therefore, more considerations should be given to individual level assignment in RD design where the sampling schemes comprise of three levels. Unviability of tools to design three-level RD studies with individual level assignment raise the question with respect to the incomplete consideration of sampling structure at the design and analysis phases. There are studies that have ignored the three-level structure at the analysis phases (Jenkins, Farkas, Duncan, Burchinal, & Vandell, 2016; Konstantopoulos & Shen, 2016; Luyten, 2006; May, Sirinides, Gray, & Goldsworthy, 2016). The issue of incomplete considerations of the sampling structure, mainly ignorance of the intermediate level, has not yet been explored in RD design and analysis literature. The only known study that has explored an incomplete consideration of sampling structure is for a clustered RE where the treatment variable resides at the top level (level 3) and concluded that ignorance of the intermediate level would not constitute a problem (Zhu, Jacob, Bloom, & Xu, 2011). There are other studies, however, that have explored this issue in depth and concluded that ignorance of an intermediate level can distort variances, estimates and their standard errors (Moerbek, 2004; Opdenakker & Van Damme, 2000; Van Den Noortgate, Opdenakker, & Onghena, 2005), a finding contradicts with Zhu, Jacob, Bloom, and Xu (2011). Although Zhu, Jacob, Bloom, and Xu (2011) has found that ignoring an intermediate level at the design phase may not have adverse consequences for cluster randomized trials, such results may
not generalize to RD studies particularly where treatment status is determined based on some assignment variable with respect to a predetermined cutoff score.

Furthermore, scholars have been increasingly designing RD studies where RE is not feasible. However, the cost associated with sampling units place constraints to achieving the desired level of precision. Considering the fact that RD designs require larger sample sizes, conducting a rigorous power analysis to determine required sample sizes has gained attention. Conventional optimal design literature in RE studies assume unlimited units are at the disposal of the researcher at each level which is hardly applicable to reality. For example, in education research for each level there might be limited units available in the population; such as the number of students in a classroom may hardly exceed 30 or the number of classroom in a school may hardly exceed five. Hedges and Borenstein (2014) proposed conditional designs where researcher can fix some of the sample sizes for some levels and find optimal sample sizes for the remaining levels. The idea of constrained optimal sample allocation first appears in Hedges and Borenstein (2014) in the education field although the rationale has been implemented in other studies (Dong & Maynard, 2013, Konstantopoulos, 2011, 2013b; Liu, 2003; Raudenbush, 1997; Raudensbush & Liu, 2000; Rhoads & Dye, 2016). The literature varies with naming the sample size calculation routines with different type of constraints, although all, one way or another, implement the notion of constrained optimal sample allocation. For example Dong and Maynard (2013) considered only sample size constraints and referred to sample size calculation as minimum required sample sizes, Hedges and Borenstein (2014) considered cost and sample size constraints and referred to
the sample size calculations *conditional optimal designs*\(^4\), and others considered only cost constraints referring to sample size calculations as *optimal design*. To this date, to my knowledge only Rhoads and Dye (2016) have extended optimal design in RE literature to two-level RD studies while only considering cost constraints. A general framework that would unify the literature under *constrained optimal sample allocation* notion is needed.

### 1.3 Significance and Contribution of the Current Study

The aforementioned concerns warrant the development of statistical power analysis tools for RD studies where individuals are the level of assignment and the sample comprise of three levels. Therefore, in this study we extend the work of Schochet (2008b, 2009) by considering various scenarios where intercepts and slopes are either constant, fixed or random. Table 1.1 describes the key characteristics of the six designs in Schochet (2008b) and five designs proposed in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Assignment</th>
<th>Intercept</th>
<th>Slope (Treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level 2</td>
<td>Level 3</td>
</tr>
<tr>
<td>BIRD2ff1</td>
<td>Level 1</td>
<td>Fixed</td>
<td>NA</td>
</tr>
<tr>
<td>BIRD2rr1</td>
<td>Level 1</td>
<td>Random</td>
<td>NA</td>
</tr>
<tr>
<td>CRD2r2</td>
<td>Level 2</td>
<td>Random</td>
<td>NA</td>
</tr>
<tr>
<td>CRD3r3</td>
<td>Level 3</td>
<td>Random</td>
<td>Random</td>
</tr>
<tr>
<td>BCRD3ff2</td>
<td>Level 2</td>
<td>Random</td>
<td>Fixed</td>
</tr>
<tr>
<td>BCRD3rr2</td>
<td>Level 2</td>
<td>Random</td>
<td>Random</td>
</tr>
<tr>
<td>BIRD3rr2rr1</td>
<td>Level 1</td>
<td>Random</td>
<td>Random</td>
</tr>
<tr>
<td>BIRD3rc2rc1</td>
<td>Level 1</td>
<td>Random</td>
<td>Random</td>
</tr>
<tr>
<td>BIRD3fc2rr1</td>
<td>Level 1</td>
<td>Random</td>
<td>Fixed</td>
</tr>
<tr>
<td>BIRD3fc2rc1</td>
<td>Level 1</td>
<td>Random</td>
<td>Fixed</td>
</tr>
<tr>
<td>BIRD3fc2fc1</td>
<td>Level 1</td>
<td>Fixed</td>
<td>Fixed</td>
</tr>
</tbody>
</table>

\(^4\) The term *constrained optimal sample allocation* appears throughout the text, but there is no formal definition.
Note. Shaded cells indicate proposed designs in this study. BIRD: Blocked individual-level regression discontinuity. CRD: Cluster-level regression discontinuity. BCRD: Blocked cluster-level regression discontinuity. Numbers in the model names refers to levels. Lower-case letters that follow numbers refers to whether the intercept and treatment slope is f for fixed, c for constant, and r for random. The last number indicates the level of treatment or assignment variable. For example, BIRD3rr2rr1: Blocked individual-level regression discontinuity where treatment variable resides at level 1, and where intercept and treatment slope are random across level 2 and level 3 units.

Furthermore, the current state of the literature in RD design requires the evaluation of incomplete consideration of the complex sampling scheme. A researcher might design an RD study by ignoring intermediate level and use formulas available for two-level designs. In such cases, the researcher might use variance parameters from two-level studies whereas the data structure comprise of three-level which is often the case in education\(^5\). In these instances, we will explore the consequences of such decisions on the sampling variance of the treatment effect estimate and statistical power.

Finally, units or clusters in the sample may be associated with certain amount of costs which limits the number of subjects or clusters to be recruited. A researcher may wish to find the sample that costs as little as possible while preserving the desired level of statistical power or MDES. In a similar vein, given a budget a researcher may wish to find out the sample size that produces maximum statistical power or minimum MDES. Constrained sample allocation (COSA) has been explored in RE under optimal design literature (e.g., Hedges & Borenstein, 2014; Konstantopoulos, 2009b, 2011, 2013b; Raudenbush, 1997; Raudenbush & Liu, 2000) and implemented in CRT-Power

---

(Borenstein, Hedges, & Rothstein, 2012), *Optimal Design* (Raudenbush et al., 2011), and *PowerUpR* (Bulus & Dong, 2017) packages. However, studies on COSA in RD studies are scarce and limited to two-level designs (Rhoades and Dye, 2016). There does not exist a consistent COSA framework that would apply to both RE and RD designs, which allows differing marginal cost per unit depending on treatment status. Therefore, we will provide a general framework that allows for COSA in RD designs that can easily be applied to corresponding RE designs as the difference underlies in the assignment mechanism. The framework allows for COSA under a variety of constraints to find optimal sample sizes (and optimal proportion of units in treatment for RE designs if marginal cost per units differs based on treatment status).
2 Literature Review

Under accurate modeling assumptions, regression discontinuity (RD) design can produce unbiased estimate of the treatment effect (Cappelleri, 1991; Goldberger, 1972a, 1972b; Rubin, 1977; Trochim & Cappelleri, 1992) comparable to that of randomized experiments (RE). Causal treatment effect estimates obtained from RD design and RE have desirable statistical properties owing to their fully known assignment mechanisms, thus their ability to isolate the treatment status from external causes. This chapter begins with introducing theoretical background on causal inference and delineates extant literature on statistical power analysis within a RD design framework and its relation to RE. It extends the review by considering complex designs where considerable nesting emerges as a result of subjects’ location or overarching organization. It follows by a review of consequences of ignoring such multilevel structure in the design and analysis phases. Finally, this chapter ends with the gap in the literature and where this study stands.

2.1 Theoretical Underpinnings

The fundamental reasoning underlying causal inference relies on conditional statements referring to an ideal world where the quantity of interest occurs as opposed to reality. Philosopher Lewis (1973) refers to this sets of would-have-been events counterfactuals. For example, assume event A occurred in reality where we observed event B as a result of the first event. Had not event A occurred, event B would not have been possible. This reasoning applies to the field of evaluation where treatment status is manipulated to claim causal effects on an outcome. However, as Rubin (2005) noted,
counterfactuals are more relevant to the events that already have occurred. In evaluation, the event occurs after treatment status is determined and subjects are exposed to the treatment, therefore, potential outcomes term is semantically coherent with the order events. In what follows we will refer to counterfactuals as potential outcomes keeping the distinction in mind.

The effect resulting from a cause is almost always relative to some other causes (Holland, 1986). For example, event A is causing event B, not some other event such as C, D or E confounders. In an experiment this is controlled through randomization. In other words, the effect of some known or unknown confounders C, D, or E that may likely cause B are evened out through randomization. Let a population of units be I, a subject \( i \) exposed to treatment (\( T_i = 1 \)) or control (\( T_i = 0 \)) may have an observed outcome \( Y_i \). This can be expressed in the following form

\[
Y_i = (1 - T_i) \cdot Y_i(0) + T_i \cdot Y_i(1) \tag{2.1}
\]

where the potential outcome for a subject exposed to treatment is \( Y_i(1) \), and \( Y_i(0) \) if they are exposed to control condition (Imbens & Lemieux, 2008). The causal effect of the treatment status, \( T_i \), is the difference between a subject’s outcome if they are exposed to treatment condition (\( T_i = 1 \))

\[
Y_i = (1 - 1) \cdot Y_i(0) + 1 \cdot Y_i(1) \tag{2.2}
\]
and if had they been in control condition \((T_i = 0)\)

\[
Y_i' = (1 - 0) \cdot Y_i(0) + 0 \cdot Y_i(1) \tag{2.3}
\]

the difference becomes

\[
Y_i(1) - Y_i(0) \tag{2.4}
\]

or vice versa. However, the fundamental problem of causal inference is that \(Y_i(1)\) and \(Y_i(0)\) cannot be observed on the same unit \(i\) at the same time (Holland, 1986). The statistical solution to overcome this is to conduct a randomized experiment (RE) by randomizing all units of \(I\) into treatment and control groups and to obtain an average treatment effect on population \(I\). Through the randomization, on average, the two groups become similar across observed and unobserved covariates that may likely affect the treatment or outcome. The concept of isolating the treatment from external factors through randomization or other means is referred to as unconfoundedness (Heckman & Hotz, 1989; Rosenbaum & Rubin, 1983). By achieving unconfoundedness, any difference between the treatment and control groups can be solely attributed to treatment status. Therefore, the average causal effect for population \(I\) is unbiased and can be derived in the form
$$E[Y(1) - Y(0)] \quad (2.5)$$

by properties of expectation

$$E[Y(1)] - E[Y(0)] \quad (2.6)$$

which can be replaced by unbiased sample means

$$\bar{Y}(1) - \bar{Y}(0) \quad (2.7)$$

However, there are cases where randomization is not feasible due to ethical reasons, and due to need-based or merit-based considerations. In these cases, treatment status may be determined based on a continuous variable, referred to as an assignment variable in the literature. Subjects may be assigned to either treatment or control conditions based on their standing on the assignment variable justified by a predetermined cutoff value. For example, treatment assignment may be manipulated on purpose to provide scholarships to students, free/reduced lunch (FRL) benefits, or compensatory reading lessons. Scholarship may be granted based on students’ grade point average, FRL benefits may be granted based on their family income level or
compensatory reading lessons may be provided based on their reading achievement scores.

In all these cases, often a threshold or cutoff value is predetermined upon which subjects are assigned to treatment condition or otherwise. In such cases, mean differences should be adjusted for non-random assignment mechanism because if the unconfoundedness assumption does not hold correlation between treatment assignment and errors is nonzero (Bennett & Lumsdaine, 1975). If the assignment mechanism is vaguely known, statistical techniques such as propensity score matching can be utilized to approximate or achieve unconfoundedness (Rosenbaum & Rubin, 1983). If the assignment mechanism is fully known, on the other hand, special regression adjustments such as regression discontinuity (RD; Thistewhite & Campbell, 1960) can be utilized. Otherwise, if special regression adjustment is ignored, non-random assignment in RD can introduce bias in the treatment effect estimate due to correlation between the assignment variable and the treatment status (Goldberger 1972a, 1972b). In RE case, due to randomization, these two variables are orthogonal to each other, that is, the correlation between them is non-existent.

In the RD case, along with $Y$ we may observe covariates such as $Z$ and $X$, and decide treatment status if they are above or below cutoff value ($Z_0$) on assignment variable $Z$. Covariate $X$ is included to provide a broader framework, and to represent common use in RD studies to improve precision, since sample sizes required in comparison to RE can be three- to four-fold times (Schochet, 2008b, 2009). That is to say, we observe $Y_i, Z_i, X_i$, and derive $T_i$ based on $Z_0$. An approximation to $Z_0$ form left and right using limit theorem, the causal treatment effect on $i$ can be expressed as
to avoid extrapolation beyond cutoff point (Imbens & Lemieux, 2008), this can be interpreted as

$$
\lim_{Z_i \uparrow Z_0} E[Y_i \mid Z_i, X_i] - \lim_{Z_i \downarrow Z_0} E[Y_i \mid Z_i, X_i]
$$

(2.8)

Similar to RE case, these two terms cannot be observed on the same unit \(i\), but considering a very narrow range of scores around \(Z_0\), assignment to \(T\) may be determined due to fluctuation in random measurement error (Boruch, 1975; Campbell & Stanley, 1963; Lee & Lemieux, 2010). Sufficiently close to cutoff \(Z_0\), conditional on \(X\) unconfoundedness (Heckman & Hotz, 1989; Rosenbaum & Rubin, 1983) assumption holds, where subject allocation to treatment and control is not affected or determined by other external factors. Therefore, the average causal effect at the cutoff value or discontinuity point can be expressed as

$$
E[Y_i(1) \mid Z_i = Z_0, X_i] - E[Y_i(0) \mid Z_i = Z_0, X_i]
$$

(2.9)

or as conditional mean differences
\[ \bar{Y}_{|Z=z_0,X}(1) - \bar{Y}_{|Z=z_0,X}(0) \]  

(2.11)

For clarity in the notation, hereafter, we will denote this difference as

\[ \bar{Y}_{t|Z=z_0,X} - \bar{Y}_{c|Z=z_0,X} \]  

(2.12)

The conditional mean differences and the associated statistical test can be obtained using analyses of covariance (ANCOVA) framework as in the following regression equation

\[ Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 (Z_i - Z_0) + r_i, \quad r_i \sim N(0, \sigma^2_{|X}) \]  

(2.13)

where \( Y_i \) is the outcome, \( T_i \) is the treatment, \( X_i \) is the covariate, \( Z_i \) is the assignment variable for subject \( i \), \( Z_0 \) is the cutoff for the assignment variable \( Z \), and \( r_i \) is random error for subject \( i \).

Conditional on covariate \( X \) the relationship between outcome variable \( Y \) and assignment variable \( Z \) can be plotted as follows

\[ \bar{Y}_{|Z=z_0,X}(1) - \bar{Y}_{|Z=z_0,X}(0) \]
The magnitude of the jump at the discontinuity point $Z_0$ represents treatment effect estimate, $\beta_1$.

There are several assumptions regarding this RD model (Hahn, Todd, Van Der Klav, 2001; Imbens & Lemieux, 2008; Trochim, 1984). First, the assignment variable and cutoff is fully known in advance. Second, there isn’t any factor affecting discontinuity at the cutoff other than treatment status. Third, conditional distribution of $Y$ given $Z$ and $X$ is continuous. Finally, conditional treatment effect is constant across the range of $Z$ conditional on $X$, in other words regression line for treatment and control groups are parallel representing each other’s counterfactuals. Any violations pertaining
the aforementioned assumptions jeopardize internal validity of the RD design due to their effect on the treatment estimate bias.

The classification of RD is structured based on whether subjects switch the treatment status (sharp versus fuzzy type) throughout the intervention, and estimation methods (parametric versus non-parametric) (Bloom, 2012; Imbens & Lemieux, 2008). If subjects are either exposed to treatment or control conditions throughout the study, it is referred as sharp designs, as opposed to, if subjects switched the treatment status at some point perhaps aware of the benefits or for some other reasons, it is referred to as fuzz design. In RD analysis, treatment effects can be estimated using parametric and non-parametric methods. Commonly used non-parametric methods are kernel and local linear regression. These two estimation methods rely on dividing subjects into bins and usually estimate treatment effect within the bins wrapping the cutoff, width of which is referred to as bandwidth.

The literature on statistical power analysis in RD design mainly focuses on parametric form of estimation with sharp type designs. Non-parametric form of estimation has many facets that factors in treatment effect bias, such as large sample requirement in the bins near the cutoff, and its sensitivity to the choice of bandwidth. Therefore, non-parametric estimation should be considered as a complementary method for validation (Bloom, 2012; Lee & Lemieux, 2010). Although the RD literature heavily focuses on non-parametric estimation methods, formulas derived from parametric estimation method provide upper bound values for MDE(S), MRSS and lower bound values for statistical power (Schochet, 2008b). Aligned with previous studies, throughout this study, a
parametric form of the regression equation will be assumed where subjects are clearly either exposed to treatment or control condition (sharp design).

To put credibility on the results, researchers often rely on statistical significance testing. However, statistical significance test for $\beta_1$ depends on the sample size, therefore, results need to be justified as to whether non-significant results are due to having a small sample size or due to failing to reject the null hypothesis of no effect. To avoid this confusion, the study should be designed to detect certain treatment effects with sufficient statistical power.

2.2 Type I and Type II Error Rates

The conditional average treatment effect $\beta_1$ may vary from sample to sample. If there were no treatment effect on the population, we would expect to see a majority of samples drawn from the population having nil values. The rest of the samples that does have non-nil values may be due to sampling errors. With infinite samples having equal sizes drawn from the population, the ratio of defective samples producing non-nil values over all samples represents the risk a researcher would tolerate to draw wrongful decisions. This is referred as Type I error, denoted by $\alpha$. Type I error can be controlled at the analysis phase (most commonly practiced rate is 5%).

If there were treatment effects on the population, on the other hand, we would expect to see a majority of samples drawn from the population having non-nil values. The rest of the samples that does not have non-nil values may be due to sampling errors. The proportion of the samples that have non-nil values over all samples is known as statistical
power. With infinite samples having equal sizes drawn from the population, the ratio of defective samples producing nil values over all samples represents the risk a researcher would tolerate to draw wrongful decisions. This is referred to as Type II error, denoted by $\beta$. Unlike $\alpha$, $\beta$ cannot be controlled at the analysis phase, and requires backwards calculations with several assumptions regarding data, before the study is conducted.

These two sources of errors and decisions made can be summarized in the following table

<table>
<thead>
<tr>
<th>Reality</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$, nil treatment effect, is true</td>
<td>Correct decision (1-$\alpha$ % of the times)</td>
</tr>
<tr>
<td>$H_0$, nil treatment effect, is false</td>
<td>Type II error ($\beta$)</td>
</tr>
</tbody>
</table>

In evaluation, particularly in statistical power analysis, the interest pertains to the correct decision detecting non-nil treatment effect. Researchers make several assumptions regarding expected treatment effects, sample sizes and other design characteristics to control the rate of Type II error rate and therefore statistical power.

While elaborating on statistical power analysis we rely on three fundamental concepts; minimum detectable effect size (MDES), minimum required sample size
(MRSS) and statistical power. Although all three are derived from the same equation, researchers may be interested in one or more of them separately.

2.3 MDE(S), MRSS and Statistical Power

Within frequentist framework statistical power analysis is conducted to ensure inferences drawn from the sample is not as a result of fluke. Therefore, prior to a study it is crucial to determine sample size with which we would be confident at a certain level that we can detect some treatment effect with some certain magnitude if there exist any. If statistical power depends on treatment effect, one can select a minimum value below which inferences are not reliable. In other words, a minimum detectable effect (MDE) can be assumed which would yield a desired level of statistical power (Bloom, 1995).

Adapting sampling variance of the treatment effect from Bloom (2006, p. 4) for complete randomized controlled trials, sampling variance of the estimator $\beta_1$ for single level RD study can be expressed as

$$\text{var}(\bar{Y}_{t|Z=Z_0,X} - \bar{Y}_{c|Z=Z_0,X}) = D \left( \frac{\sigma^2_{\hat{y}|Z=Z_0,X}}{n_t} + \frac{\sigma^2_{\hat{y}|Z=Z_0,X}}{n_c} \right)$$ (2.14)

or

$$\text{se}(\bar{Y}_{t|Z=Z_0,X} - \bar{Y}_{c|Z=Z_0,X}) = \sqrt{D \left( \frac{\sigma^2_{\hat{y}|Z=Z_0,X}}{n_t} + \frac{\sigma^2_{\hat{y}|Z=Z_0,X}}{n_c} \right)}$$ (2.15)
where $n_t$ and $n_c$ are sample sizes below and above the cutoff for treatment and control groups, $\sigma_{Y|Z=z_0,X}^2$ is conditional variance in the outcome around the cutoff, and where

$$D = \frac{1}{1 - \rho_{TZ}^2}$$

(See Bloom, 2012; Lee & Munk, 2008; Schochet, 2008b, 2009), $\rho_{TZ}^2$ is the squared correlation between assignment variable and the treatment variable. In the case of RD the correlation between assignment variable and treatment status is non-nil because treatment status is determined based on assignment variable ($\rho_{TZ} \neq 0$), therefore the term $D$ is some value greater than unity. In the case of RE, the term $D$ reaches unity because the correlation between treatment status and assignment variable is nil due to randomization ($\rho_{TZ} = 0$). Thus equation above can be restated as in Bloom (2006, p. 4)

$$se(\bar{Y}_{t|Z=z_0,X} - \bar{Y}_{c|Z=z_0,X}) = \sqrt{D \frac{\sigma_{Y|Z=z_0,X}^2}{nP(1-P)}}$$  \hspace{1cm} (2.16)

To calculate MDE, one needs to multiply a value representing the treatment effect in $t$ distribution units, referred to as multiplier (Bloom, 1995, 2006), with standard error of the treatment effect. The multiplier depends on model associated degrees of freedom, $v$, probability of Type I error rate, $\alpha$, and probability of Type II error rate, $\beta$. One-tailed multiplier with $v$ degrees of freedom can be expressed as

$$M_v = t_\alpha + t_{1-\beta}$$  \hspace{1cm} (2.17)
and two-tailed multiplier as

\[ M_v = t_{\alpha/2} + t_{1-\beta} \]  \hspace{1cm} (2.18)

where \( \alpha \) stands for Type I error and \( \beta \) stands for Type-II error as shown in Figure 2.2 (Bloom, 2006).

![Figure 2.2 Multiplier as a function of central and non-central t distributions for one-tailed test](image)

Therefore,

\[ MDE(\bar{Y}_{t|Z=Z_0,X} - \bar{Y}_{e|Z=Z_0,X}) = M_v \sqrt{\frac{D}{nP(1-P)}} \]  \hspace{1cm} (2.19)
Dividing both side of the equation by $\sigma_Y$, unconditional standard deviation of the outcome, and restating $\frac{\sigma^2_{Y|z=z_0,x}}{\sigma^2_Y}$ as $1 - R^2$, where $R^2$ is the coefficient of determination interpreted as the proportion of variance in the outcome explained by the covariate around the cutoff value. Then minimum detectable effect size takes the form

$$MDES(\bar{Y}_{t|z=z_0,x} - \bar{Y}_{c|z=z_0,x}) = M_v \sqrt{\frac{D 1-R^2}{nP(1-P)}}$$  \hspace{1cm} (2.20)

From this equation, MRSS for $n$ can easily be obtained given MDES and other model based parameters

$$n = \left(D \frac{1-R^2}{P(1-P)}\right) \left(\frac{M_v}{MDES(\bar{Y}_{t|z=z_0,x} - \bar{Y}_{c|z=z_0,x})}\right)^2$$  \hspace{1cm} (2.21)

As for calculating statistical power one needs to know the non-centrality parameter, $\gamma$.

From Figure 2.2 it is evident that the mean of [non-central] $t$ distribution at right hand side is the non-centrality parameter, $\lambda$, which takes the form

$$\lambda = \frac{MDES(\bar{Y}_{t|z=z_0,x} - \bar{Y}_{c|z=z_0,x})}{\sqrt{\frac{P(1-P)n}{D(1-R^2)}}}$$  \hspace{1cm} (2.22)
Then power for one-tailed test can be calculated from $t$ distribution as follows (Hedges & Rhoads, 2010, p. 18):

$$power = 1 - H(c(\alpha, v), v, \lambda)$$  \hspace{1cm} (2.23)

and for two-tailed

$$power = 1 - H\left(c\left(\frac{\alpha}{2}, v\right), v, \lambda\right) + H\left(-c\left(\frac{\alpha}{2}, v\right), v, \lambda\right)$$  \hspace{1cm} (2.24)

where $H$ is cumulative distribution function of noncentral $t$-distribution given the quantile $c(\alpha, v)$ or $c\left(\frac{\alpha}{2}, v\right)$, degrees of freedom $v$, and noncentrality parameter $\lambda$; and $c$ is the quantile of $t$-distribution associated with probability of $\alpha$ or $\frac{\alpha}{2}$, and degrees of freedom $v$.

### 2.4 Relative Efficiency and Statistical Power in RD Design as opposed to RE

The earliest seminal studies on the statistical power analysis in RD design were conducted by Goldberger (1972a, 1972b). Goldberger mainly focused on the comparison of efficiency in RD design as opposed to RE. Comparing the efficiency of RD design with RE using ratio of sampling variances for treatment effect estimate under two
designs, Goldberger (1972a, 1972b) found that RD design requires 2.75 times as many participants assuming a normally distributed assignment variable. Relative efficiency of RD design compared to RE (no correlation between $Z$ and $T$ in RE case) is denoted as $D$

$$D = \frac{1}{1 - \rho_{TZ}^2}$$

(2.25)

where $\rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{p(1-p)}\sigma_Z}$ (Schochet, 2008b, 2009). From the formula it can be seen that relative efficiency is mainly driven by the distribution of the assignment variable and its correlation with the treatment status (Schochet, 2008b, 2009).

Cappilleri (1991) used RE as a baseline model where 100% of the assignment variable is randomized into treatment and control conditions. He compared several variants of cutoff based designs including, conventional RD design where subjects above a cutoff receives the treatment and subjects below a cutoff constitutes the control condition or vice versa. In addition, he includes several other hybrid designs, incorporating randomization within various intervals around the cutoff, with differing probabilities (50% is used in conventional RE). The first design Capilleri (1991) considered was restricted randomized experiment (RRE) in which subjects within the cutoff interval are randomized to treatment and control conditions and the rest are discarded from the analysis. The second design is RD design where subjects below or above the cutoff are assigned to either treatment or control conditions. The third design couples RE and RD where subjects are randomized within the cutoff interval but
treatment status for the rest is determined based on cutoff. The last design is unrestricted RE where subjects are randomized to treatment and control conditions along the continuum of the assignment variable without acknowledging the cutoff. Capilleri (1991) concluded that the more subjects are randomized to treatment and control condition the more statistical power the method has, that is the four method can be ordered as RE, RD-RE, RRE, and RD based on the magnitude of their statistical power.

As long as the product of proportion of sample falling under treatment and the control remain the same, efficiency is not affected (Cappilleri, 1991), for example, by whether the split is at 30% or 70%. Based on the equation (2.4.1), Cappilleri (1991) calculated that loss in efficiency in RD design compared to RE is 2.75 at the 50% split of sample around the cutoff, 2.64 at the 40% split, 2.3 at the 30% split, 1.96 at the 20% split, and 1.52 at the 10% split assuming a normally distributed assignment variable. In cases where the assignment variable can take various distributional forms (uniform, bimodal, truncated normal), to achieve the same statistical precision as RE, in the case of 100% of the cases are allocated to either treatment or control conditions based on the cutoff, RD design sample sizes would be 3 to 4 times larger (Schochet, 2008b, 2009). On the other hand, incorporating optimal bandwidth proposed by Imbens and Kalyanaraman (2011), Deke and Dragoset (2012) concluded RD design requires 9 to 17 times as many participants as in RE mainly driven by bandwidth selection. However, bandwidth considerations are not elaborated here because they are data driven methods.

Efficiency is related to statistical power calculations as

---

*Imbens and Kalyanaraman (2009) bandwidth selection algorithm balances the tradeoff between the bias and variance, in other words minimizes the mean squared error.*
\[ \text{var}(\bar{Y}_{t|Z=Z_0,X} - \bar{Y}_{c|Z=Z_0,X}) = \left( \frac{1}{1-\rho_{TZ}^2} \right) \left( \frac{\sigma^2_{Y|Z=Z_0,X}}{n_t} + \frac{\sigma^2_{Y|Z=Z_0,X}}{n_c} \right) \] (2.26)

where sampling variance of the conditional treatment effect is inflated by a factor of \( D = \frac{1}{1-\rho_{TZ}^2} \), in comparison to RE case.

It can be seen from (2.4.1) and (24.2) that efficiency in RD design improves as treatment-to-control ratio deviates from 50% split, as opposed to RE where efficiency is maximum at the 50% split. A fifty percent split of the sample has the least relative efficiency in RD design for a normally distributed assignment variable (Schochet, 2008b, 2009). However, deviations from balanced allocation does not contribute to statistical power much, due to offsetting property of variance balance tradeoff (Schochet, 2008b, 2009). In other words, everything else being the same, by solely manipulating the treatment-to-control ratio there is a point at which the efficiency of RD design is on a par with RE. Cappilleri (1991) calculated that efficiency reaches the same level at the 34% split for RD design and 10% split for the RE.

### 2.5 Statistical Power Analysis for RD Design with Clustering

In the case clustering is present, one of the essential assumption of analysis of (co)variance is violated where it is assumed errors are independent from each other. This is the case in education where students’ outcomes are correlated with each other due to being in the same classroom, due to using similar resources, due to being taught by the
same teacher, due to being in the same school or district. Consequences of ignoring such nesting or organizational structures has been well documented (e.g., Moerbek, 2004; Opdenakker & Van Damme, 2000; Van Den Noortgate, Opdenakker, & Onghena, 2005; Zhu, Jacob, Bloom, & Xu, 2011).


There are various approaches to address clustering effects in treatment effect estimation. To meet What Works Clearinghouse (WWC; U.S. Department of Education, 2014) standards without reservation, standard errors of the treatment effect estimate in the clustered RD design case should be properly addressed by one of the accredited methods such as boot-strap, multilevel modeling (Raudenbush & Bryk, 2002) or the method proposed by Lee and Card (2008) for addressing misspecification errors. In the field of educational evaluations, resurrection of RD design methodology came at a time when multilevel modeling is ubiquitously utilized (e.g., Gamse et al., 2008; Henry, Fortner, & Thompson, 2010; Hustedt, Jung, Barnett, & Williams, 2015; Jenkins et al., 2016; Konstantopoulos & Shen, 2016; Luyten, 2006; May, Sirinides, Gray, & Goldsworthy, 2016). Schochet (2008b) provided tools to conduct statistical power analysis for designing clustered RD design under six scenarios within multilevel modeling framework (Raudenbush & Bryk, 2002). In what follows we will present these six designs proposed.
by Schochet (2008b) in the standardized form and notation as delineated in Dong and Maynard (2013).

2.5.1 Two-level Blocked Individual RD Designs with Fixed Effects (RD2ff1)

This design pertains to two-level designs where subjects are assigned to treatment and control conditions within level 2 units (sites or blocks) based on their location on an assignment variable in comparison to a cutoff score. Assume level 2 identification variable is denoted as an indicator variable $C$ and that the intervention has fixed effects across level 2 blocks. Considering an assignment variable $Z$ with cut-off $Z_0$, from which treatment variable $T$ derived, and a covariate $X$ at level 1, the treatment effect $\beta_1$ can be estimated using the following two-level regression model.

**Level 1:**

$$Y_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}X_{ij} + \beta_{3j}(Z_{ij} - Z_0) + r_{ij}$$

**Level 2:**

$$\beta_{0j} = \gamma_{00} + \sum_{j=1}^{J-1} \gamma_{0j} C_j$$

$$\beta_{1j} = \gamma_{10} + \sum_{j=1}^{J-1} \gamma_{1j} C_j$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

where $r_{ij} \sim \mathcal{N}(0, \sigma^2_{r})$, $\gamma_{0j}$ and $\gamma_{1j}$ are fixed effects associated with intercepts and slopes for $j \in \{1, 2, \ldots, J\}$. 
Given the model above minimum detectable effect size can be calculated as (Dong & Maynard, 2013; Schochet, 2008b)

\[ MDES = M_{Jn-2J-g_1} \sqrt{\frac{D(1-R_1^2)}{P(1-P)n}} \] (2.27)

from which we can obtain minimum required sample size equation

\[ J = \left( \frac{M_{Jn-2J-g_1}}{MDES} \right)^2 \left( \frac{D(1-R_1^2)}{P(1-P)n} \right) \] (2.28)

and non-centrality parameter to calculate power as

\[ \lambda = MDES \sqrt{\frac{P(1-P)n}{D(1-R_1^2)}} \] (2.29)

where \( M_{Jn-2J-g_1} = t_\alpha + t_{1-\beta} \) with \( Jn - 2J - g_1 \) degrees of freedom. \( g_1 \) is number of covariates included at level 1. \( R_1^2 \) is the proportion of variance in the outcome between level 1 other than treatment status and assignment variable. \( D = \frac{1}{1-\rho_{TZ}^2} \) and \( \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{P(1-P)\sigma_Z}} \) under normality assumption \( \rho_{TZ}^2 = \frac{\phi(\Phi^{-1}(1-P))}{\sqrt{P(1-P)}} \), for other distributional
assumptions regarding the assignment variable $Z$ see Schochet (2008b). $P$ is average treatment-to-sample ratio per block. $n$ is average number of subjects per block.

### 2.5.2 Two-level Blocked Individual RD Designs with Random Effects (RD2rr1)

This design pertains to two-level designs where subjects are assigned to treatment and control conditions within sites or schools based on their location on an assignment variable in comparison to a cutoff score. The intervention is assumed to have varying effect across blocks. Considering an assignment variable $Z$ with cut-off $Z_0$, from which treatment variable $T$ derived, and a covariate $X$ at level 1, $W$ at level 2, the treatment effect $\beta_1$ can be estimated using the following two-level regression model

**Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}X_{ij} + \beta_{3j}(Z_{ij} - Z_0) + r_{ij}$

**Level 2:**

- $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + \mu_{0j}$

- $\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + \mu_{1j}$

- $\beta_{2j} = \gamma_{20}$

- $\beta_{3j} = \gamma_{30}$

where $r_{ij} \sim N(0, \sigma^2_X) \text{ and } (\mu_{0j}, \mu_{1j}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2_X & \tau_{2T2|W} \\ \tau_{2T2|W} & \tau_{T2|W} \end{pmatrix} \right)$.

Given the model above minimum detectable effect size can be calculated as (Dong & Maynard, 2013; Schochet, 2008b)
\[ MDES = M_{J-g_2-1} \left\{ \frac{\rho_2 \omega_2 (1-R_{2T}^2)}{J} \right\} + D \left( \frac{(1-\rho_2)(1-R_{2T}^2)}{p(1-p)jn} \right) \] (2.30)

from which we can obtain minimum required sample size equation

\[ J = \left( \frac{M_{J-g_2-1}}{MDES} \right)^2 \left( \rho_2 \omega_2 (1-R_{2T}^2) \right) + D \left( \frac{(1-\rho_2)(1-R_{2T}^2)}{p(1-p)jn} \right) \] (2.31)

and non-centrality parameter to calculate power as

\[ \lambda = MDES \left\{ \frac{p(1-p)jn}{\sqrt{p(1-p)n \rho_2 \omega_2 (1-R_{2T}^2) + D(1-\rho_2)(1-R_{2T}^2)}} \right\} \] (2.32)

where \( M_{J-g_2-1} = t_\alpha + t_{1-\beta} \) with \( J - g_2 - 1 \) degrees of freedom. \( g_2 \) is number of covariates included at level 2. \( \rho_2 \) is proportion of variance in the outcome explained between level 2 clusters (sites or blocks). \( \omega_2 \) is the variance in the treatment effect between level 2 clusters (sites or blocks) standardized by outcome variance between clusters. \( R_{2T}^2 \) is proportion of variance in the treatment effect between level 2 clusters (sites or blocks) explained by level 2 covariates. \( R_{1T}^2 \) is the proportion of variance explained in the outcome by level 1 covariates other than treatment status and assignment.
variable. $D = \frac{1}{1-\rho_{TZ}}$ and $\rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{P(1-P)}\sigma_2}$ under normality assumption $\rho_{TZ}^2 = \frac{\Phi(\Phi^{-1}(1-P))}{\sqrt{P(1-P)}}$, for other distributional assumptions regarding the assignment variable $Z$ see Schochet (2008). $P$ is average treatment-to-sample ratio per block. $J$ is number of level 2 units. $n$ is average number of subjects per level 2 units.

### 2.5.3 Two-level Cluster RD Designs with Random Effects (RD2r2)

This design pertains to two-level designs where clusters are assigned to treatment and control conditions based on their location on an (mean) assignment variable in comparison to a (mean) cutoff score. Clusters are assumed to be randomly drawn from a population. Considering an assignment variable $Z$ with cut-off $Z_0$, from which treatment variable $T$ derived, and a covariate $X$ at level 1, $W$ at level 2, the treatment effect $\gamma_{01}$ can be estimated using the following two-level regression model

**Level 1:** $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$

**Level 2:** $\beta_{0j} = \gamma_{00} + \gamma_{01}T_j + \gamma_{03}(Z_j - Z_0) + \gamma_{03}W_j + \mu_{0j}$

$\beta_{1j} = \gamma_{10}$

where $r_{ij} \sim N(0, \sigma_{rj}^2)$ and $\mu_{0j} \sim N(0, \tau_{\mu j}^2)$.

Given the model above minimum detectable effect size can be calculated as (Dong & Maynard, 2013; Schochet, 2008)
\[ MDES = M_{J-g_2-2} \left( \frac{\rho_2(1-R_2^2)}{P(1-P)J} + \frac{(1-\rho_2)(1-R_1^2)}{P(1-P)n} \right) \] (2.33)

from which we can obtain minimum required sample size equation

\[ J = \left( \frac{M_{J-g_2-2}}{MDES} \right)^2 D \left( \frac{\rho_2(1-R_2^2)}{P(1-P)} + \frac{(1-\rho_2)(1-R_1^2)}{P(1-P)n} \right) \] (2.34)

and non-centrality parameter to calculate power as

\[ \lambda = MDES \frac{P(1-P)n}{\sqrt{D(n\rho_2(1-R_2^2)+(1-\rho_2)(1-R_1^2))}} \] (2.35)

where \( M_{J-g_2-1} = t_\alpha + t_{1-\beta} \) with \( J - g_2 - 1 \) degrees of freedom. \( g_2 \) is number of covariates included at level 2. \( \rho_2 \) is proportion of variance in the outcome explained between level 2 clusters. \( R_2^2 \) is proportion of variance in the outcome between level 2 covariates other than treatment status and assignment variable. \( R_1^2 \) is the proportion of variance in the outcome between level 1 covariates. \( D = \frac{1}{1-\rho_{TZ}^2} \) and \( \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{P(1-P)\sigma_Z}} \)

under normality assumption \( \rho_{TZ}^2 = \frac{\Phi(\Phi^{-1}(1-P))}{\sqrt{P(1-P)}} \), for other distributional assumptions regarding the assignment variable \( Z \) see Schochet (2008b). \( P \) is treatment-to-sample ratio. \( J \) is number of level 2 units. \( n \) is average number of subjects per level 2 units.
2.5.4 Three-level Cluster RD Designs with Random Effects (RD3r3)

This design pertains to three-level designs where clusters are assigned to treatment and control conditions based on their location on an (mean) assignment variable in comparison to a (mean) cutoff score. Clusters are assumed to be randomly drawn from a population. Considering an assignment variable $Z$ with cut-off $Z_0$, from which treatment variable $T$ derived, and a covariate $X$ at level 1, $W$ at level 2, $V$ at level 3, the treatment effect $\xi_{001}$ can be estimated using the following three-level regression model:

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk}X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + \gamma_{01k}W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k}$

Level 3: $\gamma_{00k} = \xi_{000} + \xi_{001}T_k + \xi_{002}(Z_k - Z_0) + \xi_{003}V_k + \zeta_{00k}$

$\gamma_{01k} = \xi_{010}$

$\gamma_{10k} = \xi_{100}$

where $r_{ijk} \sim N(0, \sigma_{x_i}^2)$, $\mu_{0jk} \sim N(0, \tau_{2|W}^2)$ and $\zeta_{00k} \sim N(0, \tau_{3|V}^2)$.

Given the model above minimum detectable effect size can be calculated as (Dong & Maynard, 2013; Schochet, 2008b)
\[ MDES = M_{K-g3-2} \sqrt{D \left( \frac{\rho_3(1-R_3^2)}{P(1-P)K} + \frac{\rho_2(1-R_2^2)}{P(1-P)J} + \frac{(1-\rho_2-\rho_3)(1-R_2^2)}{P(1-P)Jn} \right)} \]  

(2.36)

from which we can obtain minimum required sample size equation

\[ K = \left( \frac{M_{K-g3-2}}{MDES} \right)^2 D \left( \frac{\rho_3(1-R_3^2)}{P(1-P)} + \frac{\rho_2(1-R_2^2)}{P(1-P)J} + \frac{(1-\rho_2-\rho_3)(1-R_2^2)}{P(1-P)Jn} \right) \]  

(2.37)

and non-centrality parameter to calculate power as

\[ \lambda = MDES \frac{P(1-P)Jn}{\sqrt{D(jn\rho_3(1-R_3^2)+n\rho_2(1-R_2^2)+(1-\rho_2-\rho_3)(1-R_2^2))}} \]  

(2.38)

where \( M_{K-g3-2} = t_\alpha + t_{1-\beta} \) with \( K - g_3 - 2 \) degrees of freedom. \( g_3 \) is number of covariates included at level 3. \( \rho_3 \) is proportion of variance in the outcome explained between level 3 clusters. \( \rho_2 \) is proportion of variance in the outcome explained between level 2 clusters. \( R_3^2 \) is proportion of variance in the outcome between level 3 covariates other than treatment status and assignment variable. \( R_2^2 \) is proportion of variance in the outcome between level 2 covariates. \( R_1^2 \) is the proportion of variance in the outcome between level 1 covariates. \( D = \frac{1}{1-\rho_4^2} \) and \( \rho_4^2 = \frac{\sigma_TZ}{\sqrt{P(1-P)\sigma_Z}} \) under normality assumption

\[ \rho_4^2 = \frac{\phi(\Phi^{-1}(1-P))}{\sqrt{P(1-P)}} \], for other distributional assumptions regarding the assignment variable
Z see Schochet (2008b). $P$ is treatment-to-sample ratio. $K$ is number of level 3 units. $J$ is average number of level 2 units per level 3 units. $n$ is average number of subjects per level 2 units.

### 2.5.5 Three-level Blocked Cluster RD Designs with Fixed Effects (RD3ff2)

This design pertains to designs where level 2 clusters are assigned to treatment and control conditions within level 3 units (sites or blocks) based on their location on an (mean) assignment variable in comparison to a (mean) cutoff score. Assume level 3 identification variable is denoted as an indicator variable $S$ and that the intervention have fixed effects across level 3 blocks. Considering an assignment variable $Z$ with cutoff $Z_0$, from which treatment variable $T$ derived, and a covariate $X$ at level 1, $W$ at level 2, the treatment effect $\gamma_{01k}$ can be estimated using the following three-level regression model

**Level 1:** $Y_{ij} = \beta_{0jk} + \beta_{1jk}X_{ijk} + r_{ijk}$

**Level 2:** $\beta_{0jk} = \gamma_{00k} + \gamma_{01k}T_{jk} + \gamma_{02k}(Z_{jk} - Z_0) + \gamma_{03k}W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k}$

**Level 3:** $\gamma_{00k} = \xi_{000} + \sum_{k=1}^{K-1} \xi_{00k}S_k$  
$\gamma_{01k} = \xi_{010} + \sum_{k=1}^{K-1} \xi_{01k}S_k$  
$\gamma_{02k} = \xi_{020}$
\( \gamma_{03k} = \xi_{030} \)
\( \gamma_{10k} = \xi_{100} \)

where \( r_{ijk} \sim N(0, \sigma^2_{X}) \), \( \mu_{0jk} \sim N(0, \tau^2_{W}) \), \( \xi_{00k} \) and \( \xi_{01k} \) are fixed effects associated with intercepts and slopes for \( k \in \{1,2,\ldots,K\} \).

Given the model above minimum detectable effect size can be calculated as

\[
MDES = M_{K(J-2) - g_2} \sqrt{D \left( \frac{\rho_2(1-R_2^2)}{P(1-P)JK} + \frac{(1-\rho_2)(1-R_2^2)}{P(1-P)JKn} \right)}
\]

(2.39)

from which we can obtain minimum required sample size equation

\[
K = \left( \frac{M_{K(J-2) - g_2}}{MDES} \right)^2 D \left( \frac{\rho_2(1-R_2^2)}{P(1-P)J} + \frac{(1-\rho_2)(1-R_2^2)}{P(1-P)Jn} \right)
\]

(2.40)

and non-centrality parameter to calculate power as

\[
\lambda = MDES \frac{P(1-P)JKn}{\sqrt{D(n\rho_2(1-R_2^2) + (1-\rho_2)(1-R_2^2))}}
\]

(2.41)
where $M_{K(J-2)-g_2} = t_\alpha + t_{1-\beta}$ with $K(J-2) - g_2$ degrees of freedom. $g_2$ is number of covariates included at level 2. $\rho_2$ is proportion of variance in the outcome explained between level 2 clusters. $R_2^2$ is proportion of variance in the outcome between level 2 covariates. $R_1^2$ is the proportion of variance in the outcome between level 1 covariates.

$D = \frac{1}{1-\rho_{TZ}^2}$ and $\rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{P(1-P)\sigma_Z}}$ under normality assumption $\rho_{TZ}^2 = \frac{\phi(\Phi^{-1}(1-P))}{\sqrt{P(1-P)}}$, for other distributional assumptions regarding the assignment variable $Z$ see Schochet (2008b). $P$ is the average treatment-to-sample ratio per level 3 units. $K$ is number of level 3 units. $J$ is average number of level 2 units per level 3 units. $n$ is average number of subjects per level 2 units.

2.5.6 Three-level Blocked Cluster RD Designs with Random Effects (RD3rr2)

This design pertains to three-level designs where level 2 clusters are assigned to treatment and control conditions within level 3 units (sites or blocks) based on their location on an (mean) assignment variable in comparison to a (mean) cutoff score. The intervention is assumed to have same varying effect across blocks. Considering an assignment variable $Z$ with cut-off $Z_0$, from which treatment variable $T$ derived, and a covariate $X$ at level 1, $W$ at level 2, $V$ at level 3, the treatment effect $\gamma_{01k}$ can be estimated using the following three-level regression model

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk}X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + \gamma_{01k}T_{jk} + \gamma_{02k}(Z_{jk} - Z_0) + \gamma_{03k}W_{jk} + \mu_{0jk}$
\[ \beta_{1jk} = \gamma_{10k} \]

Level 3: \( \gamma_{00k} = \xi_{000} + \xi_{001}V_k + \zeta_{00k} \)

\[ \gamma_{01k} = \xi_{010} + \xi_{011}V_k + \zeta_{01k} \]

\[ \gamma_{02k} = \xi_{020} \]

\[ \gamma_{03k} = \xi_{030} \]

\[ \gamma_{10k} = \xi_{100} \]

where \( r_{ijk} \sim N(0, \sigma^2_{ij|X}) \), \( \mu_{0jk} \sim N(0, \tau^2_{2|W}) \) and \( \begin{pmatrix} \xi_{00k} \\ \xi_{01k} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2_{3|V} & \tau_{3|3}|V \\ \tau_{3|3}|V & \tau^2_{3|V} \end{pmatrix} \right) \).

Given the model above minimum detectable effect size can be calculated as (Dong & Maynard, 2013; Schochet, 2008b)

\[
MDES = M_{K-g_3-1} \sqrt{\frac{\rho_3 \omega_3 (1-R^2_{37})}{K}} + D \left( \frac{\rho_2 (1-R^2_2)}{P(1-P)JK} + \frac{(1-\rho_2-\rho_3)(1-R^2_2)}{P(1-P)Kn} \right) 
\]

(2.42)

from which we can obtain minimum required sample size equation

\[
K = \left( \frac{M_{K-g_3-1}}{MDES} \right)^2 \left( \rho_3 \omega_3 (1-R^2_{37}) + D \left( \frac{\rho_2 (1-R^2_2)}{P(1-P)J} + \frac{(1-\rho_2-\rho_3)(1-R^2_2)}{P(1-P)Kn} \right) \right)
\]

(2.43)
and non-centrality parameter to calculate power as

\[
\lambda = MDES \frac{P(1-P)JKn}{\sqrt{P(1-P)Jn\rho_3\omega_3(1-R_{3T}^2)+D(\rho_2(1-R_{2}^2)+(1-\rho_2-\rho_3)(1-R_{3}^2))}}
\]  

(2.44)

where \( M_{K-g_3-1} = t_{\alpha} + t_{1-\beta} \) with \( K - g_3 - 1 \) degrees of freedom. \( g_3 \) is number of covariates included at level 3. \( \rho_3 \) is proportion of variance in the outcome explained between level 3 clusters (sites or blocks). \( \rho_2 \) is proportion of variance in the outcome explained between level 2 clusters. \( \omega_3 \) is the variance in the treatment effect between level 3 clusters (sites or blocks) standardized by outcome variance between clusters. \( R_{3T}^2 \) is proportion of variance in the treatment effect between level 3 clusters (sites or blocks) explained by level 3 covariates. \( R_{2}^2 \) is the proportion of variance explained in the outcome by level 2 covariates other than treatment status and assignment variable. \( R_{1}^2 \) is the proportion of variance explained in the outcome by level 1 covariates. \( D = \frac{1}{1-\rho_T^2} \) and

\[\rho_T^2 = \frac{\sigma_{TZ}}{\sqrt{P(1-P)}} \] under normality assumption \( \rho_T^2 = \frac{\phi(\Phi^{-1}(1-P))}{\sqrt{P(1-P)}} \), for other distributional assumptions regarding the assignment variable \( Z \) see Schochet (2008b). \( P \) is average treatment-to-sample ratio per block. \( J \) is number of level 2 units. \( J \) is average number of level 2 units per level 3. \( n \) is average number of subjects per level 2 units.
2.6 Consequences of Ignoring a Level of Nesting Structure

The omission of intermediate level is common in practice, sometimes due to the absence of administrative records that identify which classroom or teacher the child belongs (Zhu, Jacob, Bloom, & Xu, 2011), or due to simplicity or small sample sizes (Van Den Noorthgate, Opdenakker, & Onghena, 2005). For the former case, the analyst has no choice but to pursue without intermediate level information, for the latter case, however, the analyst should acknowledge the intermediate level. Even if sample sizes at the intermediate level are very small, this problem can be addressed by using bootstrapping or Bayesian methods (Goldstein, 2011) or by introducing level 2 information as fixed effects into the model (Van Den Noorthgate, Opdenakker, & Onghena, 2005). Another way to decide whether to acknowledge or ignore an intermediate level is to base the modeling decision on the model fit (Opdenakker & Van Damme, 2000). If the chi-square test of difference is meaningful between the model that ignores and the model that acknowledges the intermediate level, then it is wise to acknowledge the intermediate level and pursue the analysis accordingly. If the data permits, to mitigate the problem, the least an analyst could do is to introduce predictors belonging to ignored level (Opdenakker & Van Damme, 2000).

This study however, solely focuses on the case of ignoring intermediate level despite all the conditions claim otherwise, in both the design and the analysis phase where the analyst have the option to design or analyze two- versus three-level study.

In education, the most common version of ignoring a level of nesting occurs with the ignorance of the classroom level information. The proportion of variance attributed to classroom level can exceed that of school level (Goldstein, 2011; Muthen, 1991), or the
magnitude of this variance can be subject specific. For instance, the proportion of variance in the mathematics achievement attributed to classroom level is higher than the proportion of variance in the reading achievement attributed to classroom level compared to school level (Nye, Konstantopoulos, & Hedges, 2004; Raudenbush & Bryk, 2002). Despite the possibility of sizeable proportion of variance attributed to the intermediate level, a plethora of empirical studies did not acknowledge classroom level information in the analysis (e.g., Konu, Lintonen, & Autio, 2002; Raudenbush & Bryk, 1986). Some recent evaluation studies indicate that RDD is not exempt from this fact (Jenkins et al., 2016; Konstantopoulos & Shen, 2016, Luyten, 2006; May, Sirinides, Gray, & Goldsworthy, 2016). The literature consistently demonstrated that ignoring a top or intermediate level has a detrimental effect on variance components attributed to a specific level, estimates for predictors, and their standard errors.

2.6.1 Effects of Ignoring a Level of Nesting on Variance Components and their Standard Errors

Some studies reported the effect of ignoring a level of nesting on variance components (Moerbek, 2004; Opdenakker & Van Damme, 2000), while some studies focused on both variance components and their standard errors (Van Den Noortgate, Opdenakker, & Onghena, 2005; Zhu, Jacob, Bloom, & Xu, 2011). Using a three-level model (students – classrooms – schools), Moerbek (2004) concluded that ignoring a level affects variance components. In the case of balanced design, Moerbek (2004) found that ignoring level 3 does not affect the variance component at level 1 but inflates the variance component at level 2 equal to the amount ignored. Using four-level model
(students – teachers – classrooms – schools), Van Den Noortgate, Opdenakker, and Onghena (2005) concluded that omission of level 4 does not affect variance estimates at level 2 and level 1, but the variance is reflected at a level just below, level 3.

The consequences of ignoring an intermediate level is more complicated. Van Den Noortgate, Opdenakker, and Onghena (2005) found that the omission of intermediate levels (level 2 or level 3) result in inflated variance estimates at the flanking levels. For example, if level 3 is omitted the variance is distributed to level 2 and level 4, a finding confirming Moerbek (2004) and Opdenakker and Van Damme (2000). However, Moerbek (2004) concluded, if the variance attributed to ignored level 2 is small, variance components are not affected to a great extent, a finding confirmed by using empirical elementary school dataset where variance attributed to classroom level is relatively small (Zhu, Jacob, Bloom, & Xu, 2011).

Moerbek further concluded that inflation in variance components depends on the ignored level (level 2 versus level 3), level at which predictor variable is measured, the magnitude of the variance component of the ignored level, and sample sizes at one or more level. Furthermore, Van Den Noortgate, Opdenakker, and Onghena (2005) found that ignoring top level inflates standard error of the variance estimates at the level 3. Ignoring intermediate levels inflates that standard errors of the variance estimates at an adjacent lower level but decrease standard errors of the variance estimates at the adjacent top level.
2.6.2 Effects of Ignoring a Level of Nesting on Fixed Effect Estimates and their Standard Errors

The effect of ignoring a nesting structure on fixed effect estimates and their standard errors are non-trivial. If the variance component of a given level is affected due to ignorance of a level nesting, it is very likely that fixed effect estimates or their standard errors at that level and those at the ignored level will be affected (Opdenakker & Van Damme, 2000). While estimates themselves may not be affected as much, detrimental effects were seen on standard errors.

In the case of balanced design, using a three-level model (students – classrooms – schools), Moerbek (2004) found that ignoring level 3 does not affect the intercept at level 1 and fixed effect estimates at level 2. Moerbek further concluded that inflation in standard errors of the fixed effect estimates depends on the ignored level (level 2 versus level 3), level at which predictor variable is measured, magnitude of the proportion of variance attributed to ignored level, and sample sizes at one more level (depending on the level ignored). For example, Moerbek (2004) found that ignoring level 2 inflates standard errors for the fixed effect estimates at level 1 resulting in high p-values, but not that of at level 3 in the case of balanced design. However, as Moerbek (2004) noted, if the proportion of variance attributed to ignored level 2 is small, standard errors of fixed effect estimates are not affected to a great extent, a finding later confirmed by Zhu, Jacob, Bloom, and Xu (2011) using elementary school data.

---

7 Balanced design is defined as having same number of lower level units per higher level units. For example, a balanced two-level design would have n number of level 1 units for each level 2 unit.
Using a four-level model (students – teachers - classrooms – schools), Van Den Noortgate, Opdenakker, and Onghena (2005) found that, in general, the standard error of the intercept and estimates of predictors at the ignored or adjacent levels are affected. When level 4 is ignored, the standard error of the estimate for predictors at level 3 and those belong to level 4 but included in the model as level 3 predictors. This applies to both balanced and unbalanced data. In balanced data, when level 3 is ignored, standard error for the predictors at level just below increases while the standard error of the intercept and estimates for predictors at the ignored level decreases. When the data is unbalanced, however, in addition to effects in balanced case, standard error of the estimates for predictors at level 4 decreases.

Opdenakker and Van Damme (2000) found that regardless of the level ignored, standard error of the intercept is underestimated. However, standard error of the estimates for predictors at level 1 and level 2 were not affected much when level 4 is ignored. If the predictor itself belongs to ignored level, then standard error of their estimates are underestimated. Zhu, Jacob, Bloom, and Xu (2011) extends previous work on ignoring a nesting structure in multilevel settings, by mainly focusing on the design phase rather than analysis, although results apply to both. In particular authors considered information from two-level data (design parameters) to design three-level studies. Manipulating and analyzing four empirical multisite datasets (including both elementary and secondary school data), Zhu, Jacob, Bloom, and Xu (2011) concluded that ignoring the intermediate level has no substantial effects on statistical power or precision or standard error of the estimate for predictor at level 3. Additionally, Zhu, Jacob, Bloom, and Xu (2011) concluded that using design parameters from two-level studies to design three level
studies does not create a substantial problem. This holds regardless of the magnitude of middle level variance component (based on empirical secondary school data), sample size of students per classroom and schools, and for models with or without covariates.

2.6.3 Effects of Ignoring a Level of Nesting in Unbalanced Designs compared to Balanced Designs

The principles that inflect influence on the variance components and standard error of the fixed effect estimates may not govern in the case of unbalanced designs. For example, Manatunga, Hudges, and Chen (2001) found that standard error of the fixed effect estimates are smaller when the intermediate level is unbalanced.

In the case of unbalanced design, Moerbek (2004) found that variance components are affected similar to the case of balanced design, that is, if level 3 is ignored the amount of variance is transferred to level 2, if level 2 is ignored the amount of variance is transferred to both level 1 and level 2 in the amount as a function of their sample sizes. However, what is unique to unbalanced cases is that ignoring a level of nesting can affect standard error of the fixed effect estimates at level 2 or level 3. For example, if level 2 is ignored the standard error of the estimate for the predictor at level 3 is slightly overestimated.

Van Den Noortgate, Opdenakker, and Onghena (2005) found that, in balanced data, when level 3 is ignored, standard error for the predictors at level just below increases while standard error of the intercept and estimates for predictors at the ignored level

---

8 Number of level 2 units differ for each level 3 unit.
decreases. When the data is unbalanced, however, in addition to effects in balanced case, the standard error of the estimates for predictors at level 4 decreases.

2.7 Evidence from Empirical Studies that Ignore a Level of Nesting in RD Design

There are several studies from 2000 onward with a focus on cutoff based assignment at the individual level which one way or another adjusted for clustering. About a quarter of these studies adjusted for clustering effects using hierarchical linear modeling framework (Hustedt, Jung, Barnett, & Williams, 2015; Luyten, 2006; Luyten, Peschar, & Coe, 2008; May, Sirinides, Gray, & Goldsworthy, 2016) and about a quarter of the studies used Lee and Card (2008) method (Balu, Zhu, Doolittle, Schiller, Jenkins, & Gersten, 2015; Cortes, 2015; Deke, Dragoset, Bogen, Gill, & Sekino, 2012; Harrington, Munoz, Curs, & Ehlert, 2016; Reardon, Arshan, Atteberry, & Kurlaender, 2010). The remaining studies either used bootstrap methods or none (Jenkins et al., 2016; Klerman, Olsho, & Bartlett, 2015; Leeds, McFarlin, & Duagherty, 2017; Ludwig & Miller, 2005; Matsudarie, 2008; Wong, Cook, Barnett, & Jung, 2008). The four RD studies relying on individual level cutoff-based assignment and also used hierarchical linear modeling framework are summarized as in the following. The design used in these studies is described with the terminology used in this study, and an alternative way that corresponds to one of the designs proposed in this study is stated at the end.

Hustedt, Jung, Barnett, and Williams (2015) evaluated the effectiveness of Arkansas Better Chance (ABC) initiative at kindergarten on student achievement relying on state’s strict age-based admission criteria to the program. Although Hustedt, Jung, Barnett, and Williams (2015) analyzed the data using single level analysis, district level
information is included in the model as fixed effects (corresponding to RD2fc1 design). Because within each district multiple classrooms were selected, this information could have been incorporated as an additional level. The data could have been analyzed using with fixed district effects at level 3, random classroom effects at level 2.

Luyten, Peschar, and Coe (2008) used Progress in International Reading Literacy Study (PIRLS) 2000 large scale assessment data to examine the effect of extra year of schooling on student achievement relying on the cutoff generated that split students in 9th and 10th grades. Luyten, Peschar, and Coe (2008) analyzed the data using two-level RD design where schooling effect is assumed to vary across schools (corresponding to RD2rr1 design). Alternatively, this study could have been analyzed by introducing random classroom effects, or by introducing states (or strata) within the country as fixed effects, since such information is available in the data.

Luyten (2006) used Trends in International Mathematics and Science Study (TIMSS) 1995 large scale assessment data to examine the effect of extra year of schooling on student achievement relying on the cutoff generated that split students into consecutive grades. Similar to Luyten, Peschar, and Coe (2008), a two-level RD design was used where schooling effect is assumed to vary across. Alternatively, this study could have been analyzed by introducing random classroom effects, or by introducing states (or strata) within the countries (Cyprus, Greece, England, Iceland, Japan, Norway, Scotland, and Singapore) as fixed effects, since such information is available in the data.

May, Sirinides, Gray, and Goldsworthy (2016) evaluated the effectiveness of Reading Recovery i3 Scale-Up on student’s achievement in first and third grades relying on students’ pretest score. May, Sirinides, Gray, and Goldsworthy (2016) analyzed the
data using two-level RD design where the program effect is assumed to vary across schools. Alternatively, this study could have been analyzed by introducing random classroom effects if possible, or by introducing districts as fixed effects. In summary, four RD studies relying on individual level cutoff-based assignment and also used hierarchical linear modeling framework could have been analyzed the data by acknowledging the classroom level information, or district or state level fixed effects.

2.8 Constrained Optimal Sample Allocation

Education researchers often rely on multilevel randomized experiments and quasi-experiments to draw causal inferences with respect to an intervention. Multilevel structure arises in education often due to students sharing the same teacher/classroom or school, teacher/classroom sharing the same school or location (Raudenbush & Bryk, 2002). A failure to address multilevel structure in analysis can result in downwardly biased standard errors, which leads to higher estimates of statistical power. As much as addressing multilevel structure in the analysis phase of an experiment requires advanced analytical procedures and tools to analyze the data, design of experiments under the same structure requires same level of care and rigor.

Although plethora of studies exist with respect to designing multilevel randomized experiment (RE) and regression discontinuity (RD) studies very few has focused on sample size perspective. To this point, majority of studies center their framework around statistical power and minimum detectable effect size (MDES) calculation. Furthermore, implementation of sample size calculation within multilevel modeling framework has been scarce, addressing only particular designs and scenarios.
Majority of studies has studied the sample size calculation under *optimal design* literature (Hedges & Borenstein, 2014; Konstantopoulos, 2011, 2013b; Liu, 2003; Raudenbush, 1997; Raudensbush & Liu, 2000; Rhoads & Dye, 2016), and implemented most of the ideas in CRT-Power (Borenstein, Hedges, & Rothstein, 2012) and ideas are partially implemented in Optimal Design (Raudenbush et al., 2011). Implementation has focused on cases where there is budget limitation and certain survey or treatment cost, which limits sample size calculations.

The need for optimal design emerged with the realization that higher-level units may be expensive to sample, therefore, to reconcile power deficiency researcher can resort to over-sampling of lower units (Raudenbush, 1997; Konstantopoulos, 2011; Cox & Kelcey, in preparation). In reality, lower level units are of limited quantities and researcher does not have control in most instances, which has been addressed in Hedges and Borenstein (2014). Hedges and Borenstein has proposed the idea of constrained optimal sample allocation (COSA) where sample sizes for some of the levels are fixed to specific value and the remaining levels are optimized. For example, considering education field, number of students within a classroom, or number of classrooms within a school are within a limited range which requires contrained optimal sample allocation.

In this study, constrained optimal sample allocation term is defined to be more specific and implies several aspects of the optimal design, (i) there are more than one level therefore it is an allocation problem, (ii) the allocation partially depends on some other limiting factors. Any limitation that alters resultant sample sizes for one or more level is referred as constraints. Therefore, not only a fixed budget, but also limitations with sample sizes for one or more level can impose constraints. Alternatively, researcher
may want to find most cost-efficient sample allocation, in this case, they may constrain power or minimum detectable effect size. From this framework, all the previous literature has implemented the idea of constrained optimal sample allocation to some extent. In this case, the term is broad enough to cover all previous studies, and specific enough to imply sample size calculation in multilevel experiments.

Limited versions of the COSA has been implemented in CRT-Power (Borenstein, Hedges, & Rothstein, 2012) and PowerUp! (Dong & Maynard, 2013). Different from CRT-Power COSA in PowerUp! does not accommodate a fixed budget or marginal costs, but it fixes all sample size parameters except the top-level sample size which aligns with the idea of COSA. The author has not presented the work under optimal design or COSA notion and referred to the resultant sample size as minimum required sample size (MRSS).

2.9 Gap in the Literature and Significance of the Current Study

The contribution of this study is three-fold: (i) To provide formulas for five designs in Table 2.1 where assignment in RD design is at individual level and the data structure comprise of three-levels, (ii) to demonstrate the adverse consequences of ignoring intermediate level when designing RD designs, and (iii) to provide a general framework for solving constrained optimal sample allocation (COSA) problems for six RD designs described in Schochet (2008b) and Dong and Maynard (2013) and five designs proposed in this study.
2.9.1 Gap in Statistical Power Analysis for Clustered RD Studies

One of the essential contribution to statistical power analysis in RD design was made by Schochet (2008b, 2009). Schochet provided sampling variance of the treatment effect for various designs within a hierarchical linear modeling (Raudenbush & Bryk, 2002) framework. The six types of designs pertain mainly assignment at the cluster level in two- and three-level designs, or assignment at individual level only in two-level designs. Schochet (2008b) highlighted that, for RD design, assignment at the school level would require an estimated treatment effect size of .33 or more to be justified: Therefore, RD design is more feasible where treatment status is determined based on student-level or classroom-level assignment variables.

In education designs with more than two-levels and assignment at the individual level is common (e.g., Gamse et al, 2008; Henry, Fortner, & Thompson, 2010; Hustedt, Jung, Barnett, & Williams, 2015). For example, scholarship may be granted based on a student’s grade point average, free/reduced lunch benefits may be granted based on family income level or socio-economic status, or compensatory reading lessons may be provided based on reading achievement. Even if an intermediate level (such as classroom or teacher level) is ignored, along with student and school level some higher level clustering (such as districts) may exist which are often regarded as fixed effects in the literature.

Thus, we expand the list of RD designs in Schochet (2008b) and propose consideration of the following scenarios. Each design is begins with BIRD letters standing for blocked individual regression discontinuity, it follows by number three and two letters. The number three refers to level 3 and two letters state whether intercept and
treatment slope at level 3 are random (r), fixed (f) or constant (c) correspondingly. Similarly the following number two refers to level 2 and two letters state whether intercept and treatment slope at level 2 are random (r), fixed (f) or constant (c) correspondingly. Number one at the end refers to the level where discontinuity resides.

The proposed designs are described as follows: RD3fc2fc1 is an RD design comprise of three levels where level 1 units are unit of assignment, level 2 intercepts are fixed but treatment slope is constant, and level 3 intercepts are fixed, but treatment slope is constant. RD3fc2rc1 is an RD design comprise of three levels where level 1 units are unit of assignment, level 2 intercept id fixed but treatment slopes is constant, and level 3 intercepts are random but treatment slopes are fixed. RD3ff2rr1 is an RD design comprise of three levels where level 1 units are unit of assignment, Level 2 intercepts and treatment slopes are random, and level 3 intercepts and treatment slopes are fixed. RD3rr2rr1 is an RD design comprise of three levels where level 1 units are unit of assignment, level 2 intercepts and treatment slopes are random, and level 3 intercepts and treatment slopes are random.

2.9.2 Gap in Consequences of Ignoring Intermediate- or Top-level in Clustered RD Studies

Another concern relevant to these designs is that, Schochet (2008b) provided RD designs with individual assignment only for two-level designs. The unavailability of tools to design three-level RD studies where students are assigned based on a cutoff, raise the question with respect to ignorance of data structure in the design phase. Although Zhu et al (2011) concluded that using design parameters from two-level studies to design three
level studies does not create a substantial problem drawing from four multi-site empirical elementary and secondary school datasets, scholars in school effectiveness research portrays a different picture with some overlapping results (Moerbek, 2004; Opdenakker & Van Damme, 2000; Van Der Noortgate, Opdenakker, & Onghena, 2005). Unlike Zhu, Jacob, Bloom, and Xu (2011), these scholars usually focused on analysis phase, but results pertain to design phase of the studies as well. More importantly, the previous literature has not distinguished fixed versus random effects at various levels when they examined variance components, fixed effect estimates and their standard errors. Thus, we expand previous literature by focusing on the design phase of three-level RD studies, and examine whether it is appropriate to design a two-level RD study using parameters derived from a two-level analysis whereas the data structure consist of three-levels.

2.9.3 Gap in Optimal Design

The research on the optimal design of experimental and quasi-experimental studies has been scarce and relevant to specific cases and designs. Designing experimental and quasi-experimental studies with relevance is an important aspect of casual inference in education. When higher levels are more expensive sampling lower units should be considered (Raudenbush, 1997; Konstantopoulos, 2011). Furthermore, when sampling of the treatment units are more expensive more controls units should be sampled (Cochran, 1963; Liu, 2003). Although Liu (2013) pointed to the limitations under unequal cost per treatment and control the idea has not been implemented in any statistical software. To this date, studies and their implementation in software packages has mostly assumed equal marginal costs per treatment and control units, considered
budget as a fixed entity, and focused on balanced designs that aim to detect main
treatment effects.

Furthermore, the idea of fixing some of the sample sizes and solving for the rest allows calculation of sample sizes that is relevant to the context. Although this is more efficient than optimizing for all of the sample sizes, this may result in an inefficient design if we know that sample sizes for some of the levels are within a range rather than a fixed value. An alternative would be to optimize all sample sizes but placing bound constraints instead of fixing them to a specific value.

Constrained sample allocation (COSA) has been explored in RE literature (e.g., Hedges & Borenstein, 2014; Konstantopoulos, 2009, 2011, 2013b; Raudenbush, 1997; Raudenbush & Liu, 2000) and implemented in CRT-Power (Borenstein, Hedges, & Rothstein, 2012), Optimal Design (Raudenbush et al., 2011), and PowerUpR (Bulus & Dong, 2017) packages. However, studies on COSA in RD studies is scarce and limited to two-level designs (Rhoades and Dye, 2016). Table 2.2 outlines characteristics of COSA in afore-mentioned studies. Majority of studies focused on random assignment cases, with equal marginal costs per treatment and control conditions, and considered budget as a fixed entity. Two studies considered constraints on sample sizes, however none considered bound constraints on sample sizes.

There aren’t many statistical tools available to design RD studies. To our knowledge, only known tool incorporating Schochet’s formulas is PowerUp! (Dong & Maynard, 2014), but current version does not address constrained sample allocation (COSA) problems from the perspective of this study. There is a need for a unified framework that allows COSA for a variety of designs for both RE and RD studies.
Table 2.2 *Characteristics of COSA in Existing Literature and this Study*

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment Assignment</th>
<th>Marginal cost</th>
<th>Budget</th>
<th>Power</th>
<th>Fixed constraints on Sample sizes</th>
<th>Bound constraints on Sample sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Konstantopoulos (2011)</td>
<td>Random</td>
<td>Equal</td>
<td>Fixed</td>
<td>Maximized</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Konstantopoulos (2013)</td>
<td>Random</td>
<td>Equal</td>
<td>Fixed</td>
<td>Maximized</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Dong &amp; Maynard (2013)</td>
<td>Random</td>
<td>X</td>
<td>X</td>
<td>Fixed</td>
<td>L-1 levels</td>
<td>X</td>
</tr>
<tr>
<td>Hedges &amp; Borenstein (2014)</td>
<td>Random</td>
<td>Equal</td>
<td>Fixed</td>
<td>Maximized</td>
<td>Any level(s)</td>
<td>X</td>
</tr>
<tr>
<td>Rhoades and Dye (2016)</td>
<td>Cutoff</td>
<td>Equal</td>
<td>Fixed</td>
<td>Maximized</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>This study</td>
<td>Random or Cutoff</td>
<td>Equal or Unequal</td>
<td>Fixed</td>
<td>Maximized</td>
<td>Any level(s)</td>
<td>✓</td>
</tr>
</tbody>
</table>

*Note.* L: Total number of levels.
3 Method

In this chapter, the methodology for deriving sampling variance of the treatment effect for five RD designs with characteristics stated in Table 1.1 is described in generic form. Model equations and Monte Carlo simulations for each design is described. In addition, the methodology to examine consequences of ignoring an intermediate- or top-level in designing BIRD3rr2rr1 model is outlined. Finally, a general framework for constrained optimal sample allocation (COSA) is proposed and a method for numerical optimization is presented.

3.1 Derivation of Sampling Variance

In generic matrix algebra terms a linear mixed effect model can be stated as below

\[ Y = X\beta + U\gamma + \epsilon, \]  

(3.1)

where

- \( y \) is the \((nJK \times 1)\) vector of observed outcome,
- \( X \) is a \((nJK \times p)\) fixed effects design matrix,
- \( \beta \) is a \((p \times 1)\) fixed effects coefficients,
- \( U \) is a \((nJK \times m)\) random effect design matrix,
- \( \gamma \) is a \((m \times 1)\) random effect coefficient and \( \gamma \sim N(0, G) \),
- \( \epsilon \) is error vector and \( \epsilon \sim N(0, R) \),
and where $n$ is number of level 1 units per level 2 unit, $J$ is number of level 2 units per level 3 unit, $K$ is number of level 3 units, $p$ is number of predictors, $m$ is number of random effects.

$y$ can be expressed by means of multivariate normal distribution as

$$
E[y] = E[X\beta + U\gamma + \epsilon]
$$

$$
= E[X\beta] + E[U\gamma] + E[\epsilon]
$$

$$
= X\beta,
$$

and

$$
var[y] = var[X\beta + U\gamma + \epsilon]
$$

$$
= var[X\beta] + var[U\gamma] + var[\epsilon]
$$

$$
= UGU^T + R,
$$

where $\gamma \sim N(0, G)$ and $\epsilon \sim N(0, R)$.

Let $W = UGU^T + R$, then the marginal distribution of the outcome is

$$
y \sim N(X\beta, W). \quad (3.2)
$$
The generalized least square (GLS) estimator of the fixed effects can be estimated as

\[
\hat{\beta} = \left( \sum_{k=1}^{K} \sum_{j=1}^{J} x_{jk}^T W^{-1} x_{jk} \right)^{-1} \sum_{k=1}^{K} \sum_{j=1}^{J} x_{jk}^T W^{-1} Y_{jk},
\]

(3.3)

and sampling variance of the \( \hat{\beta} \) can be estimated via generalized least squares (GLS) as

\[
\text{var}(\hat{\beta}) = \left( \sum_{k=1}^{K} \sum_{j=1}^{J} x_{jk}^T W^{-1} x_{jk} \right)^{-1},
\]

(3.4)

from which the sampling variance of the treatment effect can be retrieved from the corresponding component. To simplify derivation, assume all variables are centralized around their mean, the data is balanced (same number level 1 units within level 2 units, and same number of level 2 units within level 3 units) and there is constant variance across higher level clusters. Considering these simplifications, for a random intercept and slope model Liu (2003) has shown that coefficients at higher levels can be stated in terms of coefficients at lower level. For a three-level model assume the vector of coefficients at level 3 is \( \xi \), and the vector coefficients at level 1 is \( \beta \). In our case \( R = \sigma^2_e I \), then conditional distribution of the outcome is

\[
y|y \sim N(X\beta + UY, \sigma^2 I).
\]

(3.5)
Considering the conditional distribution in 3.5 the relation between level 2 and level 3 coefficients under afore-mentioned assumptions can be stated as

\[ \hat{\xi} = \frac{1}{JK} \sum_{k=1}^{K} \sum_{j=1}^{J} \hat{\beta}, \quad (3.6) \]

or

\[ \hat{\xi} = \frac{1}{JK} \sum_{k=1}^{K} \sum_{j=1}^{J} (X_{jk}^T X_{jk})^{-1} X_{jk}^T Y_{jk}. \quad (3.7) \]

This means conditional on the level 2 and level 3 membership, the intercept and the treatment effect is

\[ \hat{\beta}_{jk}^T = \left[ \frac{1}{n} \sum_{i=1}^{n} Y_{ijk}, \frac{1}{nP} \sum_{i=1}^{nP} Y_{ijk} - \frac{1}{n(1-P)} \sum_{i=nP+1}^{n} Y_{ijk} \right], \quad (3.8) \]

and the ordinary least square (OLS) estimator for sampling variance of the estimator is

\[ Var(\hat{\beta}_{jk} | \Omega) = \sigma^2 (X_{jk}^T X_{jk})^{-1}. \quad (3.9) \]

The marginal variance of the level 1 coefficients can be stated as
\[ \text{Var}(\hat{\beta}_{jk}) = E(\hat{\beta}_{jk}^2) - E(\hat{\beta}_{jk})^2. \] (3.10)

Equation 3.10 can be re-stated in terms of conditional variances in the form

\[ \text{Var}(\hat{\beta}_{jk}) = E\left(\text{Var}(\hat{\beta}_{jk}|\Omega)\right) + \text{Var}\left(E(\hat{\beta}_{jk}|\Omega)\right). \] (3.11)

which simplifies to

\[ \text{Var}(\hat{\beta}_{jk}) = \sigma^2(X_{jk}^T X_{jk})^{-1} + \Omega. \] (3.12)

where \( X_{jk} \) is level 1 design matrix for fixed effects conditional on level 2 and level 3 membership, and the \( \Omega \) is pooled variance-covariance structure of the coefficients. The variance of the treatment effect can be retrieved from the associated cell of the resultant matrix.

**Proof:**

\[
\text{Var}(\hat{\beta}_{jk}) = E\left(\text{Var}(\hat{\beta}_{jk}|\Omega)\right) + \text{Var}\left(E(\hat{\beta}_{jk}|\Omega)\right)
= E\left[E(\hat{\beta}_{jk}^2|\Omega) - E(\hat{\beta}_{jk}|\Omega)^2\right] + E\left[E(\hat{\beta}_{jk}|\Omega)^2\right] - \left[E\left(E(\hat{\beta}_{jk}|\Omega)\right)^2\right]
= E\left[E(\hat{\beta}_{jk}^2|\Omega)\right] - E\left[E(\hat{\beta}_{jk}|\Omega)^2\right] + E\left[E(\hat{\beta}_{jk}|\Omega)^2\right] - E(\hat{\beta}_{jk})^2
= E(\hat{\beta}_{jk}^2) - E(\hat{\beta}_{jk})^2
\]
3.2 Models and Monte Carlo Simulations

In the following the distinction between the design and analysis phase is made by the words *design* and *model*. For example X *design* implies considerations prior to data collection for the specific RD study named X, whereas X *model* implies the statistical procedure after the data collection for the specific RD study named X.

3.2.1 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 and Level 3 Intercepts and Treatment Effect are Random (BIRD3rr2rr1)

Consider a nested sampling structure consisting of three-levels, with an assignment variable $Z$ at level 1, a predetermined cut-off $Z_0$ from which treatment variable $T$ derived, a covariate $X$ at level 1, covariate $W$ at level 2, and covariate $X$ at level 1. Assume level 2 and level 3 intercepts are random, and treatment effect is random across level 2 and level 3 units, and that data is balanced, that is, $n$ level 1 units per level 2 unit, and $J$ level 2 unit per level 3 unit, and with $K$ level 3 units.

The treatment effect can be estimated using three-level hierarchical linear model (Raudenbush & Bryk, 2002). The goal is to derive a formula for sampling variance of the treatment effect in a closed form, and to validate the formula using Monte Carlo simulations. Thus, the first step is to calculate corresponding terms in the Equation 3.12 considering the relationship between coefficients at different levels in Equations 3.6 and 3.7. The next step is to use Monte Carlo simulation to validate variance of the treatment effect calculated from formula and statistical power associated with it. Simulation procedure ensures non-centrality parameter $\lambda$ belongs to the statistical test under scrutiny.
For simulation, data will be generated in top-down fashion starting from level 3 using the full model in the following sections. The data is replicated randomly 5000 times, and each replication is analyzed using hierarchical models in the following sections using PROC MIXED in SAS with default restricted maximum likelihood (REML) estimation and unstructured (UN) variance-covariance structure. Empirical standard error is calculated as standard deviation of the 5000 treatment effect estimates, and empirical power is calculated based on the proportion of replications rejecting the null with a \( p \)-value smaller than 0.05.

Average parameter values are used to calculate standard error and statistical power using proposed formula. Empirical standard errors and empirical power is compared to calculated standard error and calculated power via absolute difference (AD) and relative difference (RD)\(^9\) as

\[
AD_{power} = \text{Calculated Power} - \text{Empirical Power}, \tag{3.13}
\]

\[
RD_{power} = 100 \times AD_{power} / \text{Empirical power}, \tag{3.14}
\]

\[
AD_{SE} = SE_{eq}(\hat{\xi}_{100}) - SE_{emp}(\hat{\xi}_{100}), \tag{3.15}
\]

\[
RD_{SE} = 100 \times AD_{SE} / SE_{emp}(\hat{\xi}_{100}), \tag{3.16}
\]

where \( \hat{\xi}_{100} \) is estimated true average treatment effect based on 5000 replications.

---

\(^9\) AD and RD are same as \textit{mean bias} and \textit{mean percent bias} as empirically obtained power and standard errors are considered true parameters.
Furthermore, coverage probabilities for the standard error that is calculated from formula is established based on proportion of times the established confidence intervals include the true treatment effect. For \( i^{th} \) replication

\[
95\% \text{ CI} = \hat{\xi}_{100|i} \pm 1.96 \times SE_{eq}(\hat{\xi}_{100}).
\]  

(3.17)

95% CI coverage rate is obtained via proportion of times \( \hat{\xi}_{100} \) - the true treatment effect - falls within the interval above.

3.2.1.1 Unconditional Model

The following unconditional model is used to obtain variance parameters \( \sigma^2, \tau^2_2 \) and \( \tau^2_3 \), as defined below, which will be used to calculate various parameters along with the parameters from full model.

Level 1: \( Y_{ij} = \beta_{0jk} + r_{ijk} \)

Level 2: \( \beta_{0jk} = \gamma_{00k} + \mu_{0jk} \)

Level 3: \( \gamma_{00k} = \xi_{000} + \zeta_{00k} \),

where \( r_{ijk} \sim N(0, \sigma^2), \mu_{0jk} \sim N(0, \tau^2_2) \) and \( \zeta_{00k} \sim N(0, \tau^2_3) \).
3.2.1.2 **Treatment Only Model**

The following model is used to obtain variance parameters $\tau_{T2}^2$ and $\tau_{T3}^2$, as defined below, which will be used to calculate various parameters along with the parameters from unconditional and full model.

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk}T_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k} + \mu_{1jk}$

Level 3: $\gamma_{00k} = \xi_{00} + \zeta_{00k}$

$\gamma_{10k} = \xi_{10} + \zeta_{10k}$.

where $r_{ijk} \sim N(0, \sigma_{ij}^2), (\mu_{0jk} \quad \mu_{1jk}) \sim N\left(0, \begin{pmatrix} \tau_{T2}^2 & \tau_{T2}^2 \\ \tau_{T2}^2 & \tau_{T2}^2 \end{pmatrix} \right)$ and

$\begin{pmatrix} \zeta_{00k} \\ \zeta_{10k} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \tau_{T3}^2 & \tau_{T3}^2 \\ \tau_{T3}^2 & \tau_{T3}^2 \end{pmatrix} \right)$.

3.2.1.3 **Full Model**

The following model is used to generate the data for Monte Carlo simulation. It is also used to obtain variance parameters $\sigma_{\mid X}^2, \tau_{2\mid W}^2$, and $\tau_{3\mid V}^2$, as defined below, which are used to calculate various parameters along with the parameters from unconditional and
treatment only model. In addition to estimation of treatment effect, empirical standard error and empirical power are estimated using this model.

Level 1: \( Y_{ij} = \beta_{0jk} + \beta_{1jk}T_{ijk} + \beta_{2jk}(Z_{ijk} - Z_0) + \beta_{3jk}X_{ijk} + r_{ijk} \)

Level 2: \( \beta_{0jk} = \gamma_{00k} + \gamma_{01k}W_{jk} + \mu_{0jk} \)
\[ \beta_{1jk} = \gamma_{10k} + \gamma_{11k}W_{jk} + \mu_{1jk} \]
\[ \beta_{2jk} = \gamma_{20k} \]
\[ \beta_{3jk} = \gamma_{30k} \]

Level 3: \( \gamma_{00k} = \xi_{000} + \xi_{001}V_k + \varsigma_{00k} \)
\[ \gamma_{10k} = \xi_{100} + \xi_{101}V_k + \varsigma_{10k} \]
\[ \gamma_{20k} = \xi_{200} \]
\[ \gamma_{30k} = \xi_{300} \]
\[ \gamma_{01k} = \xi_{010} \]
\[ \gamma_{11k} = \xi_{110} \]

where \( r_{ijk} \sim N(0, \sigma^2_{[X]}, \begin{pmatrix} \mu_{0jk} \\ \mu_{1jk} \end{pmatrix}) \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2_{2|W} & \tau^2_{2|W} \\ \tau^2_{2|W} & \tau^2_{2|W} \end{pmatrix} \right) \) and

\( \begin{pmatrix} \varsigma_{00k} \\ \xi_{10k} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2_{3|V} & \tau^2_{3|V} \\ \tau^2_{3|V} & \tau^2_{3|V} \end{pmatrix} \right) \) and where
\[ \rho_2 = \frac{\tau_2^2}{\tau_3^2 + \tau_2^2 + \sigma^2} \] and represents proportion of variance in the outcome between level 2 units. \[ \rho_3 = \frac{\tau_3^2}{\tau_3^2 + \tau_2^2 + \sigma^2} \] and represents proportion of variance in the outcome between level 3 units,

\[ \omega_2 = \frac{\tau_2^2}{\tau_2^2} \] and represents treatment effect heterogeneity across level 2 units,

\[ \omega_3 = \frac{\tau_3^2}{\tau_3^2} \] and represents treatment effect heterogeneity across level 3 units,

\[ \sigma^2 \] is level 1 variance,

\[ \tau_{3|V}^2 \] is level 3 variance conditional on level 3 variables,

\[ \tau_{2|W}^2 \] is level 2 variance conditional on level 2 variables,

\[ R_1^2 = 1 - \sigma_{|X}^2 / \sigma^2 \] and is level 1 variance explained by level 1 variables,

\[ R_{2T}^2 = 1 - \frac{\tau_{2|W}^2}{\tau_{2|W}^2} \] and is proportion of variance at level 2 on the treatment explained by level 2 variables,

\[ R_{3T}^2 = 1 - \frac{\tau_{3|V}^2}{\tau_{3|V}^2} \] and is proportion of variance at level 3 on the treatment explained by level 3 variables,

\[ D = \frac{1}{1 - \rho_{TZ}} \] , \[ \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sigma_T \sqrt{P(1-P)\sigma_Z}} \] \( \sigma_{TZ} \) is covariance between \( T \) and \( Z \) and \( \sigma_Z \) is standard deviation of \( Z \).
3.2.1.4 Monte Carlo Simulation

Generate $Z, X, W, V \sim N(0,1)$ and derive $T$ from $Z$ such that $P = 0.5$ or $0.2$.

Manipulate coefficients such that $\rho_2$ and $\rho_3$ values are close to those commonly encountered in education setting. The two scenarios that produce different values of $\rho_2$ and $\rho_3$ are as follows:

Scenario 1

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 0.5 (Z_{ijk} - Z_0) + 0.5 X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + 0.3 W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k} + 0.3 W_{jk} + \mu_{1jk}$

Level 3: $\gamma_{00k} = 0 + 0.25 V_k + \zeta_{00k}$

$\gamma_{10k} = \xi_{100} + 0.25 V_k + \zeta_{10k}$,

where $r_{ijk} \sim N(0,1)$, $(\mu_{0jk} \mu_{1jk}) \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 1.5 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 \end{pmatrix}\right)$ and $(\zeta_{00k} \zeta_{10k}) \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 \end{pmatrix}\right)$.

Scenario 2

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 0.3 (Z_{ijk} - Z_0) + 0.3 X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + 0.25 W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k} + 0.25 W_{jk} + \mu_{1jk}$

Level 3: $\gamma_{00k} = 0 + 0.2 V_k + \zeta_{00k}$

$\gamma_{10k} = \xi_{100} + 0.2 V_k + \zeta_{10k}$.
where \( r_{ijk} \sim N(0,3), (\mu_{0j.k}, \mu_{1j.k}) \sim N \left( \begin{pmatrix} 0 \\ 1.5 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right) \) and \( (\varsigma_{00k}, \varsigma_{510k}) \sim N \left( \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0.5 \end{pmatrix} \right) \).

Along with the four scenarios (Scenario 1 or 2, by \( P = 0.5 \) or 0.2) above, differ treatment effect for statistical power analysis as \( \xi_{100} = 0.25 \) and for Type I error analysis as \( \xi_{100} = 0.25 \). Additionally differ sample size \( K = 50 \) or 100 and use \( n = 20 \), and \( J = 5 \) across all the scenarios. Sample sizes are chosen to approximate those commonly encountered in education. Although \( J = 5 \) may not be as common, to obtain consistent variance estimates it is an ideal minimum number. In total there are eight scenarios for statistical power analysis and there are eight scenarios for Type I error analysis.

### 3.2.2 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 and Level 3 Intercepts are Random but Treatment Effect is Constant (BIRD3rc2rc1)

Different from design BIRD3rr2rr1, treatment effect is constant across level 2 and level 3 units. Derivation focuses on the second component (\( \Omega \)) of the Equation 3.12 as this is the only difference. Similar to BIRD3rr2rr1 model data generation mechanism follows the full model, however, contrary to BIRD3rr2rr1 design, random treatment effects are not included. For this reason, to obtain design parameters only unconditional and full model are used. The form of unconditional model is same as BIRD3rr2rr1 design, and it is used to obtain variance parameters \( \sigma^2, \tau^2_2 \) and \( \tau^2_3 \) as defined previously.
3.2.2.1 Full Model

The following model is used both to generate the data for Monte Carlo simulation, and to obtain variance parameters $\sigma^2_{|X}$, $\tau^2_{2|W}$, and $\tau^2_{3|V}$, as defined below, which are used to calculate various parameters along with the parameters from unconditional model. As with BIRD3r2r1 design, treatment effect, empirical standard error and empirical power are estimated using this model.

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk}T_{ijk} + \beta_{2jk}(Z_{ijk} - Z_0) + \beta_{3jk}X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + \gamma_{01k}W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k}$

$\beta_{2jk} = \gamma_{20k}$

$\beta_{3jk} = \gamma_{30k}$

Level 3: $\gamma_{00k} = \xi_{000} + \xi_{001}V_k + \zeta_{00k}$

$\gamma_{10k} = \xi_{100}$

$\gamma_{20k} = \xi_{200}$

$\gamma_{30k} = \xi_{300}$

$\gamma_{01k} = \xi_{010}$

$\gamma_{11k} = \xi_{110}$
where \( r_{ijk} \sim N(0, \sigma_{ij}^2), \mu_{0jk} \sim N(0, \tau_{2|w}^2) \) and \( \zeta_{000k} \sim N(0, \tau_{3|v}^2) \) and where \\

\( \sigma^2 \) is level 1 variance, \\

\[ \rho_2 = \frac{\tau_2^2}{\tau_3^2 + \tau_2^2 + \sigma^2} \] and represents proportion of variance in the outcome between level 2 units, \\

\[ \rho_3 = \frac{\tau_3^2}{\tau_3^2 + \tau_2^2 + \sigma^2} \] and represents proportion of variance in the outcome between level 3 units, \\

\( \tau_{3|v}^2 \) is level 3 variance conditional on level 3 variables, \\

\( \tau_{2|w}^2 \) is level 2 variance conditional on level 2 variables, \\

\[ R_1^2 = 1 - \frac{\sigma_{ij}^2}{\sigma^2} \] and is level 1 variance explained by level 1 variables, \\

\[ R_2^2 = 1 - \frac{\tau_{2|w}^2}{\tau_2^2} \] and is level 2 variance explained by level 2 variables, \\

\[ R_3^2 = 1 - \frac{\tau_{3|v}^2}{\tau_3^2} \] and is level 3 variance explained by level 3 variables, \\

\[ D = \frac{1}{1 - \rho_{TZ}}, \quad \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{p(1-p)\sigma_Z}}, \quad \sigma_{TZ} \] is covariance between \( T \) and \( Z \) and \( \sigma_Z \) is standard deviation of \( Z \).

### 3.2.2.2 Monte Carlo Simulation

Generate \( Z, X, W, V \sim N(0, 1) \) and derive \( T \) from \( Z \) such that \( P = 0.5 \) or 0.2. Manipulate coefficients such that \( \rho_2 \) and \( \rho_3 \) values are close to those commonly
encountered in education setting. The two scenarios that produce different values of $\rho_2$ and $\rho_3$ are as follows:

**Scenario 1**

Level 1:  $Y_{ij} = \beta_{0jk} + 0.3T_{ijk} + 1(Z_{ijk} - Z_0) + 1.5X_{ijk} + \tau_{ijk}$

Level 2:  $\beta_{0jk} = \gamma_{00k} + 1W_{jk} + \mu_{ojk}$

$\beta_{1jk} = \gamma_{10k}$

Level 3:  $\gamma_{00k} = 0 + 0.75V_k + \zeta_{00k}$

$\gamma_{10k} = \xi_{100}$,

where $r_{ijk} \sim N(0,2)$, $\mu_{ojk} \sim N(0,2)$ and $\zeta_{00k} \sim N(0,1)$.

**Scenario 2**

Level 1:  $Y_{ij} = \beta_{0jk} + \beta_{1jk}T_{ijk} + 0.3(Z_{ijk} - Z_0) + 1.5X_{ijk} + \tau_{ijk}$

Level 2:  $\beta_{0jk} = \gamma_{00k} + 0.75W_{jk} + \mu_{ojk}$

$\beta_{1jk} = \gamma_{10k}$

Level 3:  $\gamma_{00k} = 0 + 0.6V_k + \zeta_{00k}$

$\gamma_{10k} = \xi_{100}$,

where $r_{ijk} \sim N(0,2)$, $\mu_{ojk} \sim N(0,1)$ and $\zeta_{00k} \sim N(0,0.5)$.
Along with the four scenarios (Scenario 1 or 2, by \( P = 0.5 \) or 0.2) above, differ treatment effect for statistical power analysis as \( \xi_{100} = 0.3 \) and for Type I error analysis as \( \xi_{100} = 0 \). Additionally differ sample size \( K = 8, 15 \) or 35 and use \( n = 30 \), and \( J = 5 \) across all the scenarios. Three different sample size scenarios are chosen to obtain a range of power values. In total there are 12 scenarios for statistical power analysis and there are 12 scenarios for Type I error analysis.

3.2.3 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts Mean Effects are Random but Treatment Effect is Random and Level 3 Mean Effects are Fixed but Treatment Effect is Constant (BIRD3fc2rr1)

Different from design BIRD3rr2rr1, level 3 intercepts are fixed and treatment effect is constant across level 3 units. Derivation focuses on the second component (\( \Omega \)) of the Equation 3.12. Level 3 fixed effects only affect \( R_2^2 \) and does not change derivation for the first component of the Equation 3.12. Similar to BIRD3rr2rr1 model data generation mechanism follows the full model, however contrary to BIRD3rr2rr1 design, random treatment effects are not included and level 3 effects are included in the model as fixed effects. To obtain design parameters, a two-level unconditional, treatment only and full models are required.
3.2.3.1 **Unconditional Model**

The form of the unconditional model is similar to BIRD3rr2rr1 design, however, level 3 intercepts are fixed. Fixed effect are included in the full model as another covariate at level 2 therefore unconditional model in this case consist of two levels. This model is used to obtain variance parameters for $\sigma^2$ and $\tau^2_2$ only as defined previously.

Level 1: $Y_{ij} = \beta_{0jk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + \mu_{0jk}$,

where $r_{ijk} \sim N(0, \sigma^2)$, $\mu_{0jk} \sim N(0, \tau^2_2)$.

3.2.3.2 **Treatment Only Model**

The following model is used to obtain variance parameters $\tau^2_T$ as defined below, which will be used to calculate various parameters along with the parameters from unconditional and full model.

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k} + \mu_{1jk}$. 
where $r_{ijk} \sim N(0, \sigma^2_{\text{r}})$, and $(\mu_{0jk}, \mu_{1jk}) \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2_{2|T} & \tau^{2\text{r} 2}_{1|T} \\ \tau^{2\text{r} 2}_{2|T} & \tau^{2\text{r} 2}_{2|T} \end{pmatrix} \right)$.

### 3.2.3.3 Full Model

As with the unconditional model and treatment only model the following full model also consist of two levels. Data generation for Monte Carlo simulation follows this model, and is to obtain variance parameters $\sigma^2_{\text{r}}$, and $\tau^2_{2|w}$, as defined below, which are used to calculate various parameters along with the parameters from unconditional and treatment only models. As with BIRD3rr2rr1 design, treatment effect, empirical standard error and empirical power are estimated using this model.

**Level 1:** $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + \beta_{2jk} (Z_{ijk} - Z_0) + \beta_{3jk} X_{ijk} + r_{ijk}$

**Level 2:** $\beta_{0jk} = \gamma_{00k} + \gamma_{01k} W_{jk} + \mu_{0jk}$

\[ \beta_{1jk} = \gamma_{10k} + \gamma_{11k} W_{jk} + \mu_{1jk} \]

\[ \beta_{2jk} = \gamma_{20k} \]

\[ \beta_{3jk} = \gamma_{30k} \]

**Level 3:** $\gamma_{00k} = \xi_{000} + \sum_{k=1}^{K-1} \xi_{00k}$

\[ \gamma_{10k} = \xi_{100} \]

\[ \gamma_{20k} = \xi_{200} \]
\[ \gamma_{30k} = \xi_{300} \]
\[ \gamma_{01k} = \xi_{010} \]
\[ \gamma_{11k} = \xi_{110}, \]

where \( r_{ijk} \sim N(0, \sigma^2_{i|x}) \), \( \left( \mu_{0j|k}, \mu_{1j|k} \right) \sim N\left( (0, 0), \begin{pmatrix} \tau^2_{2|W} & \tau^2_{2|W} \\ \tau^2_{2|W} & \tau^2_{2|W} \end{pmatrix} \right) \), and \( \xi_{00k} \) are fixed effects associated with level 3 means for \( k \in \{1, 2, ..., K - 1\} \) constrained to have mean of zero, and where

\[ \sigma^2 \] is level 1 variance,
\[ \tau^2_2 \] is level 2 variance,
\[ \tau^2_{2|W} \] is level 2 variance conditional on level 2 variables,

\[ \omega_2 = \frac{\tau^2_2}{\tau^2_2} \] and represents treatment effect heterogeneity across level 2 units,

\[ \rho_2 = \frac{\tau^2_2}{\tau^2_2 + \sigma^2} \] and represents proportion of variance in the outcome between level 2 units,

\[ R^2_1 = 1 - \sigma^2_{i|x}/\sigma^2 \] and is level 1 variance explained by level 1 variables,

\[ R^2_{2T} = 1 - \tau^2_{2|W}/\tau^2_{2T} \] and is proportion of variance at level 2 on the treatment explained by level 2 variables,

\[ D = \frac{1}{1 - \rho^2_{TZ}}, \quad \rho^2_{TZ} = \frac{\sigma_{TZ}}{\sqrt{p(1-p)\sigma^2_Z}}, \quad \sigma_{TZ} \text{ is covariance between } T \text{ and } Z \text{ and } \sigma^2_Z \text{ is standard deviation of } Z. \]
Monte Carlo Simulation

Generate $Z, X, W \sim N(0,1)$ and derive $T$ from $Z$ such that $P = 0.5$ or 0.2. Manipulate coefficients such that $\rho^2$ values are close to those commonly encountered in education setting. The two scenarios that produce different values of $\rho^2$ are as follows:

Scenario 1

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 1(Z_{ijk} - Z_0) + 1X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + 0.5W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k} + 0.5W_{jk} + \mu_{1jk}$

Level 3: $\gamma_{00k} = 0 + \sum_{k=1}^{K-1} \xi_{00k}$

$\gamma_{10k} = \xi_{100}$

where $r_{ijk} \sim N(0,2), \begin{pmatrix} \mu_{0jk} \\ \mu_{1jk} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \right)$ and fixed effects $\xi_{00k} \sim N(0,1)$.

Scenario 2

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 2(Z_{ijk} - Z_0) + 2X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + 0.25W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k} + 0.25W_{jk} + \mu_{1jk}$

Level 3: $\gamma_{00k} = 0 + \sum_{k=1}^{K-1} \xi_{00k}$

$\gamma_{10k} = \xi_{100}$.
where $r_{ijk} \sim N(0, 2), \begin{pmatrix} \mu_{0jk} \\ \mu_{1jk} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix} \right)$ and fixed effects $\xi_{00k} \sim N(0, 1)$.

Along with the four scenarios (Scenario 1 or 2, by $P = 0.5$ or 0.2) above, differ treatment effect for statistical power analysis as $\xi_{100} = 0.35$ and for Type I error analysis as $\xi_{100} = 0$. Additionally differ sample size $K = 10, 20$ or 40 and use $n = 20$, and $J = 5$ across all the scenarios. In total there are 12 scenarios for statistical power analysis and there are 12 scenarios for Type I error analysis.

3.2.4 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts are Random but Level 3 Intercepts are Fixed, and Treatment Effect is Constant across Level 2 and Level 3 (BIRD3fc2rc1)

Different from design BIRD3rr2rr1, level 3 intercepts are fixed and treatment effect is constant across level 2 and level 3 units. Derivation focuses on the second component ($\Omega$) of the Equation 3.12. Similar to BIRD3fc2rr1 level 3 fixed effects are included at level 2 and factors in through $R_{2}^2$. As with previous designs, data generation mechanism follows the full model. For this reason, to obtain design parameters only unconditional and full model are used. The form of unconditional model is same as BIRD3fc2rc1 design, and it is used to obtain variance parameters $\sigma^2$, and $\tau_{2}^2$ as defined previously.
3.2.4.1  **Full Model**

The following model is used both to generate the data for Monte Carlo simulation, and to obtain variance parameters $\sigma^2_{X_2}$, and $\tau^2_{W_2}$ as defined below, which are used to calculate various parameters along with the parameters from unconditional model. As with BIRD3rr2rr1 design, treatment effect, empirical standard error and empirical power are estimated using this model.

Level 1:  $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + \beta_{2jk} (Z_{ijk} - Z_0) + \beta_{3jk} X_{ijk} + r_{ijk}$

Level 2:  $\beta_{0jk} = \gamma_{00k} + \gamma_{01k} W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k}$

$\beta_{2jk} = \gamma_{20k}$

$\beta_{3jk} = \gamma_{30k}$

Level 3:  $\gamma_{00k} = \xi_{000} + \sum_{k=1}^{K-1} \xi_{00k}$

$\gamma_{10k} = \xi_{100}$

$\gamma_{20k} = \xi_{200}$

$\gamma_{30k} = \xi_{300}$

$\gamma_{01k} = \xi_{010}$

$\gamma_{11k} = \xi_{110}$,
where \( r_{ijk} \sim N(0, \sigma_{ijk}^2) \), \( \mu_{0jk} \sim N(0, \tau_{2|W}^2) \), and \( \xi_{00k} \) are fixed effects associated with level 3 means for \( k \in \{1,2, ..., K\} \) constrained to have mean of zero, and where \( \sigma^2 \) is level 1 variance,

\[
\rho_2 = \frac{\tau_T^2}{\tau_T^2 + \sigma^2}
\]
and represents proportion of variance in the outcome between level 2 units,

\[
R_1^2 = 1 - \frac{\sigma_{|X}^2}{\tau_T^2} \text{ and is level 1 variance explained by level 1 variables,}
\]

\[
D = \frac{1}{1-P_{TZ}} \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{P(1-P)\sigma_Z}}, \text{ } \sigma_{TZ} \text{ is covariance between } T \text{ and } Z \text{ and } \sigma_Z \text{ is standard deviation of } Z.
\]

### 3.2.4.2 Monte Carlo Simulation

Generate \( Z, X, W \sim N(0,1) \) and derive \( T \) from \( Z \) such that \( P = 0.5 \) or \( 0.2 \). Manipulate coefficients such that \( \rho_2 \) values are close to those commonly encountered in education setting. The two scenarios that produce different values of \( \rho_2 \) are as follows:

**Scenario 1**

Level 1: \( Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 1(Z_{ijk} - Z_0) + 1X_{ijk} + r_{ijk} \)

Level 2: \( \beta_{0jk} = \gamma_{00k} + 0.6W_{jk} + \mu_{0jk} \)

\[
\beta_{1jk} = \gamma_{10k}
\]

Level 3: \( \gamma_{00k} = 0 + \sum_{k=1}^{K-1} \xi_{00k} \)

\[
\gamma_{10k} = \xi_{100},
\]
where $r_{ijk} \sim N(0,1.5)$, $\mu_{0jk} \sim N(0,0.6)$, and fixed effects $\xi_{00k} \sim N(0,0.1)$.

**Scenario 2**

Level 1: $Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 0.25 (Z_{ijk} - Z_0) + 0.5 X_{ijk} + r_{ijk}$

Level 2: $\beta_{0jk} = \gamma_{00k} + 0.6 W_{jk} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k}$

Level 3: $\gamma_{00k} = 0 + \sum_{k=1}^{K-1} \xi_{00k}$

$\gamma_{10k} = \xi_{100}.$

where $r_{ijk} \sim N(0,1.5)$, $\mu_{0jk} \sim N(0,0.6)$, and fixed effects $\xi_{00k} \sim N(0,0.1)$.

Along with the four scenarios (Scenario 1 or 2, by $P = 0.5$ or $0.2$) above, differ treatment effect for statistical power analysis as $\xi_{100} = 0.25$ and for Type I error analysis as $\xi_{100} = 0$. Additionally differ sample size $K = 10, 20$ or $40$ and use $n = 20$, and $J = 5$ across all the scenarios. In total there are 12 scenarios for statistical power analysis and there are 12 scenarios for Type I error analysis.

### 3.2.5 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts and Level 3 Intercepts are Fixed and Treatment Effect is Constant (BIRD3fc2fc1)

Different from design BIRD3rr2rr1, level 2 and level 3 intercepts are fixed and treatment effect is constant across level 2 and level 3 units. Derivation focuses on the second component ($\Omega$) of the Equation 3.12. Level 2 and level 3 fixed effects are
included at level 1 and factors in through $R_1^2$. As with previous designs data generation mechanism follows the full model. Unconditional model is not needed as variance of the outcome ($\sigma^2$) represents variance obtained from unconditional model. For this reason, to obtain design parameters only the full model is used.

### 3.2.5.1 Full Model

The following model is used both to generate the data for Monte Carlo simulation, and to obtain variance parameters $\sigma_{\bar{x}}^2$ as defined below, which is used to calculate $R_2^2$ value along with $\sigma^2$. As with BIRD3rr2rr1 design, treatment effect, empirical standard error and empirical power are estimated using this model.

**Level 1:**

$$Y_{ij} = \beta_{0jk} + \beta_{1jk}T_{ijk} + \beta_{2jk}(Z_{ijk} - Z_0) + \beta_{3jk}X_{ijk} + r_{ijk}$$

**Level 2:**

$$\beta_{0jk} = \gamma_{00k} + \sum_{j=1}^{K-1} \gamma_{0jk}$$

$$\beta_{1jk} = \gamma_{10k}$$

$$\beta_{2jk} = \gamma_{20k}$$

$$\beta_{3jk} = \gamma_{30k}$$

**Level 3:**

$$\gamma_{00k} = \xi_{000} + \sum_{k=1}^{K-1} \xi_{00k}$$

$$\gamma_{10k} = \xi_{100}$$

$$\gamma_{20k} = \xi_{200}$$
\[ \gamma_{30k} = \xi_{300} \]
\[ \gamma_{01k} = \xi_{010} \]
\[ \gamma_{11k} = \xi_{110}, \]

where \( r_{ijk} \sim N(0, \sigma_{\alpha}^2) \), \( \mu_{0jk} \) and \( \xi_{00k} \) are fixed effects associated with level 2 and level 3 means for \( jk \in \{1,2,\ldots,JK\} \) and \( k \in \{1,2,\ldots,K\} \) constrained to have mean of zero, and where

\( \sigma^2 \) is level 1 variance,

\[ R_1^2 = 1 - \frac{\sigma_{\alpha}^2}{\sigma^2} \]

and is level 1 variance explained by level 1 variables,

\[ D = \frac{1}{1 - \rho_{TZ}^2}, \quad \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{\sigma_T \sigma_Z}}, \quad \sigma_{TZ} \text{ is covariance between } T \text{ and } Z \text{ and } \sigma_Z \text{ is standard deviation of } Z. \]

### 3.2.5.2 Monte Carlo Simulation

Generate \( Z, X \sim N(0,1) \) and derive \( T \) from \( Z \) such that \( P = 0.5 \) or 0.2. Manipulate coefficients such that \( R_1^2 \) have different values representing high and low exploratory power of the covariates (e.g., high exploratory power present when pretest is included in a model). The two scenarios that produce different values of \( R_1^2 \) are as follows:

**Scenario 1**

Level 1: \[ Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 1(Z_{ijk} - Z_0) + 1X_{ijk} + r_{ijk} \]

Level 2: \[ \beta_{0jk} = \gamma_{00k} + \sum_{j=1}^{JK} \gamma_{0jk} \]
\[ \beta_{1jk} = \gamma_{10k} \]

Level 3: \[ \gamma_{00k} = 0 + \sum_{k=1}^{K-1} \xi_{00k} \]

\[ \gamma_{10k} = \xi_{100}. \]

where \( r_{ijk} \sim N(0,1) \), fixed effects \( \gamma_{0jk} \sim N(0,1) \), and fixed effects \( \xi_{00k} \sim N(0,1) \).

Scenario 2

Level 1: \[ Y_{ij} = \beta_{0jk} + \beta_{1jk} T_{ijk} + 0.25(Z_{ijk} - Z_0) + 0.25X_{ijk} + r_{ijk} \]

Level 2: \[ \beta_{0jk} = \gamma_{00k} + \sum_{j,k=1}^{JK-1} \gamma_{0jk} \]

\[ \beta_{1jk} = \gamma_{10k} \]

Level 3: \[ \gamma_{00k} = 0 + \sum_{k=1}^{K-1} \xi_{00k} \]

\[ \gamma_{10k} = \xi_{100}. \]

where \( r_{ijk} \sim N(0,1.5) \), fixed effects \( \gamma_{0jk} \sim N(0,1) \), and fixed effects \( \xi_{00k} \sim N(0,1) \).

Along with the four scenarios (Scenario 1 or 2, by \( P = 0.5 \) or 0.2) above, differ treatment effect for statistical power analysis as \( \xi_{100} = 0.25 \) and for Type I error analysis as \( \xi_{100} = 0 \). Additionally differ sample size \( K = 10 \) or 25 and use \( n = 20 \), and \( J = 5 \) across all the scenarios. In total there are eight scenarios for statistical power analysis and there are eight scenarios for Type I error analysis.
3.3 Consequences of Ignoring Intermediate- or Top-Level

This section outlines the methodology to examine consequences of ignoring intermediate- or top-level in BIRD3rr2rr1 design on statistical power analysis. When intermediate- or top-level is ignored, the BIRD3rr2rr1 design becomes a BIRD2rr1 design (Dong & Maynard, 2013; Schochet, 2008b). Due to unavailability of tools for three-level studies, a researcher may be tempted to ignore an intermediate- or top-level and design an RD study using BIRD2rr1 design as opposed to BIRD3rr2rr1 using parameters from two-level studies that already ignored intermediate- or top-level. This might be the case when intermediate- or top-level information had been missing in the data. Parameters obtained from these misspecified models are distorted due to ignorance of a level. Calculated sample sizes would be incorrect had these distorted parameters been used in the power analysis to design a two-level study.

BIRD3rr2rr1 design is of interest merely due to availability of corresponding two-level BIRD2rr1 design and due to its prevalent use among education scholars. The data is already generated and analyzed using full model in BIRD3rr2rr1 design. In addition, a two-level unconditional, treatment only, and full models are fitted to the same data by either ignoring the intermediate- or top-level. When intermediate level is ignored in BIRD3rr2rr1 model, for example, schools become level 2 units in the new misspecified BIRD2rr1 model. When top level is ignored, classrooms become level 2 in the new misspecified BIRD2rr1 model. For a two-level misspecified BIRD2rr1 model the unstandardized form of the variance of the treatment effect is

\[
Var(\hat{\tau}_{100}) = \frac{\tau^2_{\hat{Y}_{|W}}}{K} + \frac{D\sigma^2_X}{nKP(1-P)},
\]

(3.18)
and in the standardized form

\[
\frac{\text{Var}(\xi_{100})}{\tau_2^2 + \sigma^2} = \frac{\rho_2 \omega_2 (1 - R_{2T}^2)}{JK} + \frac{D(1 - \rho_2)(1 - R_1^2)}{nJKP(1 - P)},
\]

(3.19)

where

\[\rho_2 = \frac{\tau_2^2}{\tau_2^2 + \sigma^2}\] and represents proportion of variance in the outcome between level 2 units

(\sigma^2 \text{ and } \tau_2^2 \text{ are obtained from unconditional model}),

\[\omega_2 = \frac{\tau_{T2}^2}{\tau_2^2}\] and represents treatment effect heterogeneity across level 2 units (\(\tau_{T2}^2\) is obtained from treatment only model, and \(\tau_2^2\) is obtained from unconditional model),

\[\sigma_{|X}^2\] is level 1 variance conditional on level 1 variables (obtained from full model),

\[\tau_{2|W}^2\] is level 2 variance conditional on level 2 variables (obtained from full model),

\[R_1^2 = 1 - \sigma_{|X}^2 / \sigma^2\] and is level 1 variance explained by level 1 variables,

\[R_{2T}^2 = 1 - \tau_{T2|W}^2 / \tau_{T2}^2\] and is proportion of variance at level 2 on the treatment explained by level 2 variables,

\[D = \frac{1}{1 - \rho_{TZ}^2},\]

\[\rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{P(1 - P)\sigma_Z}}.\]
\( \sigma_{TZ} \) is covariance between \( T \) and \( Z \),

and \( \sigma_Z \) is standard deviation of \( Z \).

There are two possible scenarios where a researcher may conduct a flawed
power analysis: (i) a researcher may use design parameters from a correctly specified
three-level model to calculate statistical power for a two-level model. In this case
researcher may either ignore intermediate-level and use variance components of level 3 in
Equation 3.19 with a top-level sample size \( K \), or ignore top-level and use variance
components of level 2 in Equation 3.19 with top-level sample size \( JK \), (ii) a researcher
may also use design parameters from a misspecified two-level model where either
intermediate- or top-level is ignored to calculate power for a two-level model. In this
case, parameters are included in Equation 3.19 and therefore the statistical power is
incorrect. The distinction between the two scenarios is that, in the former, parameters
used in the BIRD2rr1 design are from the correctly specified BIRD3rr2rr1 model(s),
whereas, in the latter, parameters used in BIRD2rr1 design are from the misspecified
BIRD2rr1 model(s).

### 3.4 Constrained Optimal Sample Allocation

Parameters that have certain limitations and influence sample size allocation are
considered to have constraints. The constraints may be due to budget limitations where
each unit is associated with certain cost or due to sampling units having limited range in
the population. To this date, among the studies that considered budget and sample size
constraints, the majority have focused on randomized experiments. In this section, we
propose a general framework for constrained optimal sample allocation (COSA) in randomized experiments and regression discontinuity studies.

### 3.4.1 Framework

For a flexible COSA, three form of constraints are proposed; primary (explicit) constraints, secondary (implicit) constraints, and tertiary (required) constraints. The primary constraints are explicit and can be placed on either total cost, power, or minimum detectable effect size (MDES). The secondary constraints are implicit and takes effect once sample sizes for one or more levels are specified in the model (point constrained or fixed). The tertiary constraints are required by default, and are on bounds, which depends on minimum degrees of freedom for the design, but if not, bounds range from zero and infinity and can be altered.

COSA problems can be solved in the following forms,

i. under fixed budgetary constraints given marginal costs per treatment and control units while minimizing sampling variance of the treatment effect

ii. under fixed power or MDES constraints given marginal costs per treatment and control units while minimizing total cost and

iii. under fixed sample size constraints or under bound constraints for one or more levels along with any of the i, or ii options.

Let $N$ be a vector of sample sizes for $L$ number of levels in a hierarchical structure, consisting of $n_1, n_2, ..., n_L$, let also $p$ be proportion of units in the treatment group, $\delta$, the estimand of interest. For the sake of brevity the other design parameters will be omitted
from the equations. The sampling variance of the treatment effect, \( \delta \), will be stated as a function of \( N \) and \( p \) as

\[
\text{var}(\delta|N, p) = f(N, p).
\] (3.20)

Then non-centrality parameter is

\[
\lambda = \frac{\delta}{\sqrt{f(N, p)}}.
\] (3.21)

Assuming a continuous outcome \( Y \) has a variance of \( \sigma_Y^2 \) then minimum detectable effect size (MDES) given a specific power is

\[
\text{MDES}(N, p|\text{Power}) = \lambda \sqrt{f(N, p)/\sigma_Y^2}.
\] (3.22)

Then power of test given a specific MDES can be calculated from \( t \) distribution as follows (Hedges & Rhoads, 2010, p. 18)

\[
\text{Power}(N, p|\text{MDES}) = \begin{cases} 
1 - H(c(\alpha, v), v, \lambda), & \text{one tailed} \\
1 - H \left( c \left( \frac{\alpha}{2}, v \right), v, \lambda \right) + H \left( -c \left( \frac{\alpha}{2}, v \right), v, \lambda \right), & \text{two tailed}
\end{cases}
\] (3.23)

where \( H \) is cumulative distribution function of noncentral \( t \)-distribution given the quantile \( c(\alpha, v) \) or \( c \left( \frac{\alpha}{2}, v \right) \), degrees of freedom \( v \), and noncentraility parameter \( \lambda \); and \( c \) is the quantile of \( t \)-distribution associated with probability of \( \alpha \) or \( \frac{\alpha}{2} \), and degrees of freedom \( v \).
Assuming sampling of each unit comes with certain marginal cost the total cost is

\[ g(N, p) = p \left( \sum_{j=1}^{i} c_{tj} \prod_{i=j}^{L} n_i \right) + (1 - p) \left( \sum_{j=1}^{i} c_{cj} \prod_{i=j}^{L} n_i \right), \]

where \( c_{tj} \) is marginal cost per treatment unit for \( j^{th} \) level, and \( c_{cj} \) is marginal cost per control unit for \( j^{th} \) level, \( n_i \) is the average sample size for \( i^{th} \) level. Unequal costs only applies to randomization level or below, and the rest of the levels are forced to have equal costs per unit. These equations constitutes the core of the framework. Combination of these produce result in two distinct routines each answering a particular research question.

### 3.4.1.1 Primary Constraint on Cost

To answer the question “Given marginal costs per units and a fixed budget, what is the optimal allocation of subjects/clusters across levels to achieve highest level of precision?”

Minimize sampling variance of the treatment

\[
\min_{N, p \in \mathbb{R}^+ \text{ and } 0 < p < 1} f(N, p). \tag{3.25}
\]

subject to primary equality constraint

\[ g(N, p) = \text{Budget}, \tag{3.26} \]
where \( N, p \in \mathbb{R}^+ \) and \( 0 < p < 1 \).

### 3.4.1.2 Primary Constraint on Power or MDES

Alternatively, to answer the question “Given marginal costs per units, what is the most cost-efficient allocation of subjects/clusters across levels given certain level of power or MDES?”

Minimize total cost

\[
\min_{N,p \in \mathbb{R}^+ \text{ and } 0 < p < 1} g(N,p),
\]

subject to primary equality constraint

\[
\text{Power}(N,p|MDES) = \text{Power},
\]

or

\[
\text{MDES}(N,p|\text{Power}) = \text{MDES},
\]

where \( N, p \in \mathbb{R}^+ \) and \( 0 < p < 1 \).

### 3.4.1.3 Secondary and Tertiary Constraints on Sample Sizes

Along with primary constraints, secondary and tertiary constraints can be placed on sample sizes. Secondary constraints can be placed on sample sizes by defining single
value for a level, in this case, \( \exists N = \exists A \), where \( A = \{a_1, a_2, ..., a_L\} \) and \( N, A \in \mathbb{R}^+ \).

Alternatively tertiary constraints can be placed on one or more level by re-defining bounds, in this case, \( LB < \exists N < UB \), where \( LB, UN \in \mathbb{R}^+ \).

It is possible for user-defined sample sizes and lower bounds to violate the minimum degrees of freedom requirement in various designs. In this case, sample sizes are adjusted and replaced with next smallest integer value that satisfy the condition. When bound contraints are defined, the starting value is modified so experimenting various starting values under such condition would not produce expected results.

### 3.4.1.4 Numerical Optimization

\( f(N, p) \) and \( g(N, p) \) equations can be combined into single form \( h(N, p) \). We need the gradient of the function \( h(N, p) \) to find \( N \) and \( p \) such that it minimizes \( h(N, p) \)

\[
\begin{bmatrix}
\frac{\partial h(n_1,n_2,...,n_L,p)}{\partial n_1} \\
\frac{\partial h(n_1,n_2,...,n_L,p)}{\partial n_2} \\
\vdots \\
\frac{\partial h(n_1,n_2,...,n_L,p)}{\partial n_L} \\
\frac{\partial h(n_1,n_2,...,n_L,p)}{\partial p}
\end{bmatrix}
= 0.
\]

(3.30)

Solving of \( L+1 \) system of equations with \( L + 1 \) unknown, and by rejecting and penalizing any solution that does not satisfy secondary and tertiary constraints, optimal \( n_1, n_2, ..., n_L \) and \( p \) can be found.

The minimization problem can be solved with algorithms that allows non-linear optimization with non-linear constraints and preferably that has global convergence.
properties. PowerUpR uses four algorithms available in NLOPT library (Johnson, n.d.) for this purpose via nloptr (Ypma, 2014) package in R (R Core Team, 2017). More specifically, Augmented Lagrangian method is used for global optimization (AUGLAG, Birgin & Martines, 2008; Conn, Gould, & Toint, 1991) in conjunction with one of the following local optimization algorithms: Constrained Optimization by Linear Approximations (COBYLA, Powell, 1994), Low Storage BFGS (LBFGS, Liu & Nocedal, 1989), Method of Moving Asymptotes (MMA, Svanberg, 2002), or Sequential Least-Squares Quadratic Programming (SLSQP, Kraft, 1988).

Among these algorithms COBYLA is a derivative-free algorithm therefore it does not require a gradient. The rest of the algorithms are gradient based and uses numerical approximations with a quasi-newton method. These algorithms essentially combine non-linear objective function with non-linear constraint and penalizes any solution that is out of bounds. Results might differ, and global solution is not guaranteed, therefore using four algorithms together increases chances of finding a global solution. In that case, most algorithms would agree. It is recommended that users use different starting values and placing as many constraints on sample sizes as possible.

PowerUpR also uses integer approximations to search the grid using brute force. Users can change the default algorithm to the one that produced most efficient allocation, and request integer approximations. PowerUpR will search and find best integer solutions and users can decide whether there exist a better solution than the best performing algorithm. If a better solution is found, starting values and sample size constraints can be modified to obtain more precise values with the best performing algorithm.
4 Results

This chapter provides results for mainly three sections: (i) derivation and validation of formulas for sampling variance of the treatment effect in various three-level regression discontinuity (RD) studies where treatment variable resides at level 1, (ii) consequences of ignoring intermediate- and top-level when designing the most general and commonly used three-level RD design (BIRD3rr2rr1), and (iii) constrained optimal sample allocation (COSA) under budget and sample size constraints.

The first section addresses five models as described in Table 1.1. Among the five models, BIRD3rr2rr1 is the most general, for which once the sampling variance of the treatment effect estimate is derived in the closed form, the rest complies the same conditions while subject to various variance constraints. Monte Carlo simulation results indicate close correspondence between standard error, statistical power, and Type I error rates obtained empirically over 5000 replications, and the standard error, statistical power and Type I error rates calculated using the formula. Ninety-five percent CI coverage probabilities are within expected range and are hovering around 95%.

The next section compares two designs: the BIRD3rr2rr1 design proposed in this study and the BIRD2rr1 designs described in Schochet (2008b) and Dong and Maynard (2013). When intermediate- or top-level is ignored in a BIRD3rr2rr1 model, sampling variance of the treatment effect in the new BIRD2rr1 model is distorted proportional to the magnitude of the variance between the ignored levels. Ignoring intermediate level results in slightly inflated standard errors consequently statistical power is underestimated. However, ignoring intermediate level is not as problematic as ignoring the top level. When top-level is ignored, the standard errors are deflated significantly and
consequently power is overestimated to a great extent. This difference occurs despite variance between level 3 units is smaller than the variance between level 2 units.

Finally, the last section provides an application of COSA in a three-level RD study where discontinuity resides at level 3. The example considers both budget and sample size constraints. The framework and optimization procedure for COSA is implemented in PowerUpR (Bulus, & Dong, 2017) package.

4.1 Derivation and Validation of Formulas

4.1.1 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts and Level 3 Intercepts and Treatment Effect are Random (BIRD3rr2rr1)

Researcher in the literature typically analyze three-level regression discontinuity studies where the treatment variable is at level 1, and where level 2 and level 3 intercepts are random, and treatment effect varies randomly across level 2 and level 3 units. Whether the effect of assignment variable (Z) is random across level 2 and level 3 clusters bears no influence on the sampling variance of the treatment effect. Therefore in the following model assignment variable is assumed to be fixed. In this case the sampling variance of the treatment is proportional to (un)conditional residual variance at level 1, level 2 and level 3, and inversely proportional to sample sizes at each level, variance of the treatment variable and the strength of the relationship between treatment variable and assignment variable. The unstandardized form of the sampling variance of the treatment effect is

\[
Var(\hat{\xi}_{100}) = \frac{\tau_{T3|V}^2}{K} + \frac{\tau_{T2|W}^2}{JK} + \frac{D\sigma_{ZX}^2}{nJKP(1-P)},
\]  

(4.1)
and in the standardized form

\[
\frac{\text{Var}(\hat{\xi}_{100})}{\tau_2^2 + \tau_3^2 + \sigma^2} = \frac{\rho_2 \omega_3 (1-R_3^2)}{K} + \frac{\rho_2 \omega_2 (1-R_2^2)}{JK} + \frac{D(1-\rho_2-\rho_3)(1-R_3^2)}{nJKP(1-P)}, \tag{4.2}
\]

where

\[
\rho_2 = \frac{\tau_2^2}{\tau_2^2 + \tau_3^2 + \sigma^2} \text{ and represents proportion of variance in the outcome explained by level 2 units,}
\]

\[
\rho_3 = \frac{\tau_3^2}{\tau_2^2 + \tau_3^2 + \sigma^2} \text{ and represents proportion of variance in the outcome explained by level 3 units,}
\]

\[
\omega_2 = \frac{\tau_2^2}{\tau_2^2} \text{ and represents treatment effect heterogeneity across level 2 units,}
\]

\[
\omega_3 = \frac{\tau_3^2}{\tau_3^2} \text{ and represents treatment effect heterogeneity across level 3 units,}
\]

\[
\sigma^2 \text{ is level 1 variance from unconditional model,}
\]

\[
\tau_2^2 \text{ is level 2 variance from unconditional model,}
\]

\[
\tau_3^2 \text{ is level 3 variance from unconditional model,}
\]

\[
\sigma_{|X}^2 \text{ is level 1 variance conditional on level 1 variables,}
\]

\[
\tau_3^2_{|V} \text{ is level 3 variance conditional on level 3 variables,}
\]

\[
\tau_2^2_{|W} \text{ is level 2 variance conditional on level 2 variables,}
\]
\[ R_1^2 = 1 - \frac{\sigma^{2}_{I/X}}{\sigma^2} \] and is level 1 variance explained by level 1 variables,

\[ R_2^2 = 1 - \frac{\tau^{2}_{I|W}}{\tau^2_{I}} \] and is proportion of variance at level 2 on the treatment explained by level 2 variables,

\[ R_3^2 = 1 - \frac{\tau^{2}_{I|V}}{\tau^2_{I}} \] and is proportion of variance at level 3 on the treatment explained by level 3 variables,

\[ D = \frac{1}{1 - \rho^2_{TZ}} \] represents efficiency loss in comparison to a randomized experiment with the same model but no correlation between the assignment variable and treatment status,

\[ \rho^2_{TZ} = \frac{\sigma_{TZ}}{\sqrt{P(1-P)\sigma^2_Z}} \] where \( \sigma_{TZ} \) is covariance between \( T \) and \( Z \) and \( \sigma_Z \) is standard deviation of \( Z \).

The hypothesis test for the treatment effect is conducted using a non-central \( t \)-distribution with non-centrality parameter \( \frac{\hat{\xi}_{100}}{\sqrt{Var(\hat{\xi}_{100})}} \) with degrees of freedom \( K - g_3 - 1 \).

Monte Carlo simulation results for BIRD3rr2rr1 design are presented in Table 4.1 and 4.2. Results indicate that there is close correspondence between \( SE_{emp}(\hat{\xi}_{100}) \), the empirical standard error for treatment effect, and \( SE_{eq}(\hat{\xi}_{100}) \), the calculated standard error. Empirical standard errors are calculated from the standard deviation of the treatment effect estimates over 5000 replications, whereas calculated standard errors are derived from Equation 4.1 using average parameter estimates over 5000 replications. Similarly, results indicate that there is close correspondence between empirical power and calculated power values (See Table 4.1). However, the relative difference between empirical power and calculated power values tends to increase as top-level sample size become smaller. As for comparison of empirical Type I error and
calculated Type I error rates, they are similar to comparison of empirical power and
calculated power values (See Table 4.2). The only difference between the two
comparisons is that the true treatment effect is specified as zero to obtain empirical Type
I error and calculated Type I error rates. Finally, 95% confidence intervals (CI) are
established using standard error calculated via formula based on average parameter
estimates over 5000 replications and coverage probabilities are obtained based on
proportion of times the true treatment effect fall within the interval. Ninety-five percent
CI formula coverage rates hover around the expected 95%.
Table 4.1  Power Analysis for BIRD3rr2rr1 Design

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{\xi}_{100}$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$SE(\bar{\xi}_{100})$</td>
<td>0.14</td>
<td>0.18</td>
<td>0.10</td>
<td>0.13</td>
<td>0.14</td>
<td>0.19</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>$ES(\bar{\xi}_{100})$</td>
<td>0.10</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
<td>0.12</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.38</td>
<td>0.15</td>
<td>0.38</td>
<td>0.15</td>
<td>0.36</td>
<td>0.13</td>
<td>0.36</td>
<td>0.13</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>0.23</td>
<td>0.09</td>
<td>0.23</td>
<td>0.09</td>
<td>0.23</td>
<td>0.09</td>
<td>0.23</td>
<td>0.09</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.77</td>
<td>0.57</td>
<td>0.77</td>
<td>0.56</td>
<td>0.90</td>
<td>0.64</td>
<td>0.91</td>
<td>0.65</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.47</td>
<td>0.47</td>
<td>0.46</td>
<td>0.46</td>
<td>0.54</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>$R^2_T$</td>
<td>0.53</td>
<td>0.07</td>
<td>0.54</td>
<td>0.07</td>
<td>0.48</td>
<td>0.05</td>
<td>0.48</td>
<td>0.05</td>
</tr>
<tr>
<td>$R^2_{2T}$</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>$R^2_{3T}$</td>
<td>0.13</td>
<td>0.11</td>
<td>0.11</td>
<td>0.09</td>
<td>0.13</td>
<td>0.14</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$\rho_{TZ}$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$K$</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$SE_{emp}(\bar{\xi}_{100})$</td>
<td>0.14</td>
<td>0.18</td>
<td>0.10</td>
<td>0.13</td>
<td>0.14</td>
<td>0.20</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>$SE_{eq}(\bar{\xi}_{100})$</td>
<td>0.13</td>
<td>0.18</td>
<td>0.10</td>
<td>0.13</td>
<td>0.14</td>
<td>0.19</td>
<td>0.10</td>
<td>0.13</td>
</tr>
<tr>
<td>AD in SEs</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RD in SEs (%)</td>
<td>-1.44</td>
<td>0.99</td>
<td>-2.27</td>
<td>3.03</td>
<td>-2.78</td>
<td>-3.85</td>
<td>-3.28</td>
<td>-2.33</td>
</tr>
<tr>
<td>95% CI Coverage</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Empirical Power</td>
<td>0.44</td>
<td>0.30</td>
<td>0.74</td>
<td>0.52</td>
<td>0.45</td>
<td>0.26</td>
<td>0.72</td>
<td>0.45</td>
</tr>
<tr>
<td>Calculated Power</td>
<td>0.42</td>
<td>0.26</td>
<td>0.73</td>
<td>0.48</td>
<td>0.44</td>
<td>0.25</td>
<td>0.72</td>
<td>0.45</td>
</tr>
<tr>
<td>AD in Powers</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RD in Powers (%)</td>
<td>-5.19</td>
<td>-13.03</td>
<td>-0.88</td>
<td>-7.13</td>
<td>-2.03</td>
<td>-3.95</td>
<td>0.18</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Note. Results are based on 5000 replications. $\bar{\xi}_{100}$: Treatment effect. SE: Standard Error. ES: Effect size. $\rho_2$: Proportion of variance in the outcome between level 2 units. $\rho_3$: Proportion of variance in the outcome between level 3 units. $\omega_2$: Treatment effect heterogeneity across level 2 units. $\omega_3$: Treatment effect heterogeneity across level 3 units. $R^2_T$: Proportion of variance in the outcome explained level 1 covariates. $R^2_{2T}$: Proportion of variance in the treatment effect explained level 2 covariates. $R^2_{3T}$: Proportion of variance in the treatment effect explained level 3 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $\rho_{TZ}$: Correlation between the assignment variable and the treatment status. $n$: Average number of level 1 units per level 2 units, which is set to 20. $J$: Average number of level 2 units per level 3 units, which is set to 5. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
Table 4.2 Type I Error Analysis for BIRD3rr2rr1 Design

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\xi}_{100} )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( SE(\hat{\xi}_{100}) )</td>
<td>0.14</td>
<td>0.18</td>
<td>0.10</td>
<td>0.13</td>
<td>0.14</td>
<td>0.19</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>( ES(\hat{\xi}_{100}) )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \rho_2 )</td>
<td>0.39</td>
<td>0.15</td>
<td>0.39</td>
<td>0.15</td>
<td>0.36</td>
<td>0.13</td>
<td>0.36</td>
<td>0.13</td>
</tr>
<tr>
<td>( \rho_3 )</td>
<td>0.23</td>
<td>0.10</td>
<td>0.23</td>
<td>0.10</td>
<td>0.24</td>
<td>0.09</td>
<td>0.24</td>
<td>0.09</td>
</tr>
<tr>
<td>( \omega_2 )</td>
<td>0.77</td>
<td>0.57</td>
<td>0.77</td>
<td>0.56</td>
<td>0.90</td>
<td>0.64</td>
<td>0.91</td>
<td>0.65</td>
</tr>
<tr>
<td>( \omega_3 )</td>
<td>0.47</td>
<td>0.47</td>
<td>0.46</td>
<td>0.46</td>
<td>0.54</td>
<td>0.52</td>
<td>0.53</td>
<td>0.52</td>
</tr>
<tr>
<td>( R_{1}^{2} )</td>
<td>0.51</td>
<td>0.06</td>
<td>0.51</td>
<td>0.06</td>
<td>0.46</td>
<td>0.05</td>
<td>0.46</td>
<td>0.05</td>
</tr>
<tr>
<td>( R_{2}^{2} )</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>( R_{3}^{2} )</td>
<td>0.13</td>
<td>0.10</td>
<td>0.11</td>
<td>0.09</td>
<td>0.13</td>
<td>0.12</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>( \rho_{T2} )</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>( K )</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

| SE_{emp}(\hat{\xi}_{100}) | 0.14 | 0.18 | 0.10 | 0.13 | 0.14 | 0.20 | 0.10 | 0.14 |
| SE_{eq}(\hat{\xi}_{100}) | 0.13 | 0.18 | 0.10 | 0.13 | 0.14 | 0.19 | 0.10 | 0.13 |
| AD in SEs | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 |
| RD in SEs (%) | -1.10 | 1.88 | -1.23 | 2.05 | -1.70 | -3.56 | -1.62 | -2.85 |
| 95% CI Coverage | 0.95 | 0.96 | 0.95 | 0.96 | 0.95 | 0.95 | 0.95 | 0.95 |
| Empirical Type I Error | 0.06 | 0.06 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.05 |
| Calculated Type I Error | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| AD in Type I Errors | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 |
| RD in Type I Errors (%) | -13.15 | -10.71 | -2.70 | -5.26 | -8.75 | -12.54 | -8.76 | -5.70 |

Note. Results are based on 5000 replications. \( \hat{\xi}_{100} \): Treatment effect. SE: Standard Error. ES: Effect size. \( \rho_2 \): Proportion of variance in the outcome between level 2 units. \( \rho_3 \): Proportion of variance in the outcome between level 3 units. \( \omega_2 \): Treatment effect heterogeneity across level 2 units. \( \omega_3 \): Treatment effect heterogeneity across level 3 units. \( R_{1}^{2} \): Proportion of variance in the outcome explained level 1 covariates. \( R_{2}^{2} \): Proportion of variance in the treatment effect explained level 2 covariates. \( R_{3}^{2} \): Proportion of variance in the treatment effect explained level 3 covariates. \( \rho \): Proportion of variance in the treatment effect explained level 2 covariates. \( \rho_{T2} \): Correlation between the assignment variable and the treatment status. n: Average number of level 1 units per level 2 units, which is set to 20. J: Average number of level 2 units per level 3 units, which is set to 5. K: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
4.1.1.1 Derivation

Assume the following design matrix for level 1 with dimensions \( n \times 3 \) for a given school \( k \) and classroom \( j \), including only the intercept, the treatment variable and the assignment variable

\[
X_{jk} = \begin{bmatrix}
1_{nP} & (1 - P)1_{nP} & (Z - Z_0)_{nP} \\
1_{n(1-P)} & -P1_{n(1-P)} & (Z - Z_0)_{n(1-P)}
\end{bmatrix},
\]

where \( n \) is marginal number of subjects at level 1, \( P \) is marginal proportion of subjects in treatment group at level 1, all the elements of \( 1_{nP} \) and \( 1_{n(1-P)} \) are consist of integer 1, and \( Z \) and \( Z_0 \) are assignment variable and the cutoff accordingly. Furthermore, assume the treatment variable is centered around the \( P \) and assignment variable is centered around \( Z_0 \). For the sake of simplicity covariates are excluded from the design matrix, but this does not change derivations below. The vector of level 1 coefficients can be written as

\[
\boldsymbol{\beta}_{jk}^T = [\beta_{0,jk} \beta_{1,jk} \beta_{2,jk}].
\]

Liu (2003) has derived ordinary least square (OLS) estimator of the treatment effect for a two-level multisite trial from generalized least square estimator (GLS) by assuming that conditional on level 2 units there would be no design effect, therefore, treatment effect can be estimated via OLS. Assuming constant variance across clusters and a balanced data at level 2 or level 3, the average treatment effect is the average of the OLS estimators at level 1, and treatment
effect varies across higher level units. In other words, there are two variations that contribute to
the variance of an estimator and can be written as

$$\text{Var} \left( \hat{\beta}_{jk} \right) = E \left( \text{Var} \left( \hat{\beta}_{jk} \mid \Omega \right) \right) + \text{Var} \left( E \left( \hat{\beta}_{jk} \mid \Omega \right) \right), \quad (4.3)$$

where $\Omega$ is the variance-covariance matrix that is associated with the design effect

$$\Omega = \begin{bmatrix}
\omega_{11} & \omega_{12} & \omega_{13} \\
\omega_{12} & \omega_{22} & \omega_{23} \\
\omega_{13} & \omega_{23} & \omega_{33}
\end{bmatrix},$$

and where $\omega_{11}, \omega_{22}, \omega_{33}$ are the variance associated with the intercept, treatment effect, and
assignment variable that is attributed to design effect.

Conditional on the design effect $\text{Var}(\hat{\beta}_{jk} \mid \Omega) = \sigma^2 (X_{jk}^T X_{jk})^{-1}$ and $\text{Var} \left( E \left( \hat{\beta}_{jk} \mid \Omega \right) \right) = \Omega$, therefore

$$\hat{\beta}_{jk} \sim i.i.d. N \left( \beta_{jk}, \sigma^2 (X_{jk}^T X_{jk})^{-1} + \Omega \right)$$

and

$$\hat{\xi} \sim i.i.d. N \left( \xi, \frac{\sigma^2 (X_{jk}^T X_{jk})^{-1} + \Omega}{JK} \right).$$
In our case

\[ \mathbf{X}_j^T \mathbf{X}_j = \begin{bmatrix} n & \sum T_i & \sum Z_i \\ \sum T_i & \sum T_i^2 & \sum Z_i T_i \\ \sum Z_i & \sum Z_i T_i & \sum Z_i^2 \end{bmatrix} = n \begin{bmatrix} 1 & 0 & 0 \\ 0 & E(T^2) & E(ZT) \\ 0 & E(ZT) & E(Z^2) \end{bmatrix}. \]

As \( n \) approaches to infinity \( (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \) is

\[ \frac{1}{n} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{E(Z^2)}{E(T^2)E(Z^2)-(E(ZT))^2} & \frac{E(ZT)}{E(T^2)E(Z^2)-(E(ZT))^2} \\ 0 & \frac{E(ZT)}{E(T^2)E(Z^2)-(E(ZT))^2} & \frac{E(T^2)}{E(T^2)E(Z^2)-(E(ZT))^2} \end{bmatrix}. \]

Since above matrix is assumed to be asymptotically unbiased variance-covariance estimators of the population, then

\[ (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \xrightarrow{p} \frac{1}{n} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sigma_Z^2}{\sigma_T^2 \sigma_Z^2 - \sigma_T Z^2} & \frac{\sigma_T Z}{\sigma_T^2 \sigma_Z^2 - \sigma_T Z^2} \\ 0 & \frac{\sigma_T Z}{\sigma_T^2 \sigma_Z^2 - \sigma_T Z^2} & \frac{\sigma_T^2}{\sigma_T^2 \sigma_Z^2 - \sigma_T Z^2} \end{bmatrix}. \]

The cell \( (\mathbf{X}_j^T \mathbf{X}_j)^{-1}_{[2,2]} \) is associated with the treatment effect. Focusing on the treatment effect
\[ \text{Var}(\hat{\xi}_{100}) = E \left( \text{Var}(\hat{\xi}_{100}|\omega_{22}) \right) + \text{Var} \left( E(\hat{\xi}_{100}|\omega_{22}) \right), \quad (4.4) \]

or

\[ \text{Var}(\hat{\xi}_{100}) = \frac{1}{jk} \sigma^2 \left( X^T_{jk} X_{jk} \right)^{-1}_{[2,2]} + \omega_{22}. \quad (4.5) \]

Since \( \sigma^2_T = P(1 - P) \) and \( \sigma^2_{TZ} = P(1 - P)\sigma^2_Z\rho_{TZ}^2 \) (Schochet, 2008, p B.2)

\[ \sigma^2 \left( X^T_{jk} X_{jk} \right)^{-1}_{[2,2]} = \frac{\sigma^2 \sigma^2_Z}{nP(1-P)\sigma^2_Z - \rho_{TT}^2 \sigma^2_Z P(1-P)} = \frac{\sigma^2}{nP(1-P)(1-\rho_{TT}^2)}. \]

Then the equation becomes

\[ \text{Var}(\hat{\xi}_{100}) = \frac{\sigma^2}{nJP(1-P)(1-\rho_{TT}^2)} + \omega_{22}. \quad (4.6) \]

As for the second component of the equation 4.1.1.2-2 variance of the level 1 coefficients can be calculated. Vector of level 1 coefficients \( \mathbf{\beta}_{jk} \) is defined as
\[ \beta_{jk} = \begin{pmatrix} \xi_{000} + \mu_{0jk} + \zeta_{00k} \\ \xi_{100} + \mu_{1jk} + \zeta_{10k} \\ \xi_{200} \end{pmatrix}. \]

and

\[ \text{Var}(\beta_{jk}) = E(\beta_{jk}\beta_{jk}^T) = E(\beta_{jk})E(\beta_{jk})^T. \quad (4.7) \]

The second component of the Equation 4.4 can be written as

\[ \text{Var}(E(\hat{\xi}|\Omega)) = \frac{1}{JK} \text{Var}\left(\sum_{j=1}^{J} \sum_{k=1}^{K} \hat{\beta}_{jk}|\Omega\right) \]

\[ = \frac{1}{JK^2} \text{Var}\left(\sum_{j=1}^{J} \sum_{k=1}^{K} \beta_{jk}|\Omega\right). \]

Schools are independent from each other, therefore

\[ \text{Var}(E(\hat{\xi}|\Omega)) = \frac{1}{JK^2} \sum_{k=1}^{K} \text{Var}(\sum_{j=1}^{J} \beta_{jk}|\Omega). \quad (4.8) \]

Focusing on only treatment effect

121
\[ \omega_{22} = \frac{1}{J^2 K^2} \sum_{k=1}^{K} Var(\sum_{j=1}^{J} \beta_{1jk}) \]
\[ = \frac{1}{J^2 K^2} \sum_{k=1}^{K} (J^2 \tau_{T3}^2 + J \tau_{T2}^2) \]
\[ = \frac{\tau_{T3}^2}{K} + \frac{\tau_{T2}^2}{JK}. \]

The complete \( \Omega \) is

\[ Var(E(\hat{\xi}|\Omega)) = \Omega = \begin{bmatrix}
\frac{\tau_{22}^2}{JK} & \frac{\tau_{32}^2}{JK} & \frac{\tau_{23}^2}{JK} & \frac{\tau_{33}^2}{JK} & 0 \\
\frac{\tau_{22}^2}{JK} & \frac{\tau_{32}^2}{JK} & \frac{\tau_{23}^2}{JK} & \frac{\tau_{33}^2}{JK} & 0 \\
\frac{\tau_{22}^2}{JK} & \frac{\tau_{32}^2}{JK} & \frac{\tau_{23}^2}{JK} & \frac{\tau_{33}^2}{JK} & 0 \\
\frac{\tau_{22}^2}{JK} & \frac{\tau_{32}^2}{JK} & \frac{\tau_{23}^2}{JK} & \frac{\tau_{33}^2}{JK} & 0 \\
\frac{\tau_{22}^2}{JK} & \frac{\tau_{32}^2}{JK} & \frac{\tau_{23}^2}{JK} & \frac{\tau_{33}^2}{JK} & 0 \\
\end{bmatrix}. \]

Recalling that

\[ Var(\hat{\beta}_{jk}) = E(Var(\hat{\beta}_{jk}|\Omega)) + Var(E(\hat{\beta}_{jk}|\Omega)). \quad (4.9) \]

Then

\[ Var(\hat{\xi}_{100}) = E(Var(\hat{\xi}_{100}|\omega_{22})) + Var(E(\hat{\xi}_{100}|\omega_{22})). \quad (4.10) \]
So

\[ \text{Var}(\hat{\xi}_{100}) = \frac{\tau_{T3|V}^2}{K} + \frac{\tau_{T2|W}^2}{JK} + \frac{\sigma_{X}^2}{nJKP(1-P)(1-\rho_{2T}^2)}. \] (4.11)

Considering that \( R_1^2 = 1 - \frac{\sigma_X^2}{\sigma^2} \), \( R_2^2 = 1 - \frac{\tau_{T2|W}^2}{\tau_{T2}^2} \), \( R_3^2 = 1 - \frac{\tau_{T3|V}^2}{\tau_{T3}^2} \), \( \rho_2 = \frac{\tau_3^2}{\tau_2^2 + \sigma^2} \), \( \rho_3 = \frac{\tau_3^2}{\tau_3^2 + \tau_2^2 + \sigma^2} \) and by standardizing the equality by \( \tau_3^2 + \tau_2^2 + \sigma^2 \) the Equation 4.11 becomes

\[ \text{Var}(\hat{\xi}_{100}) = \frac{\rho_3 \omega_3 (1-R_{3T}^2)}{K} + \frac{\rho_2 \omega_2 (1-R_{2T}^2)}{JK} + \frac{(1-P)(1-\rho_3)}{nJKP(1-P)(1-\rho_{2T}^2)}. \] (4.12)
4.1.2 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts and Level 3 Intercepts are Random (RD3rr2rr-1) but Treatment Effect is Constant (BIRD3rc2rc1)

Researcher may want to design a three-level regression discontinuity study where the treatment variable is at level 1, and where level 2 and level 3 intercepts vary randomly, and treatment effect is constant across level 2 and level 3 units. In this case, the sampling variance of the treatment is proportional to (un)conditional residual variance, and inversely proportional to sample sizes at each level, variance of the treatment variable and the strength of the relationship between the treatment variable and the assignment variable.

Since treatment effect is constant across level 2 and level 3 units, compared to BIRD3rr2rr1 model, the terms $\omega_2 = 0$, $\omega_3 = 0$, $R_{2T}^2 = 1$, and $R_{3T}^2 = 1$. The unstandardized form of the sampling variance of the treatment effect becomes

$$Var(\hat{\xi}_{100}) = \frac{D\sigma_{\hat{\xi}}^2}{njKP(1-P)}, \quad (4.13)$$

and the standardized form is

$$Var(\hat{\xi}_{100}) = \frac{D(1-\rho_2-\rho_3)(1-R_{11}^2)}{njKP(1-P)}, \quad (4.14)$$
where

$\sigma^2$ is level 1 variance from unconditional model,

$\tau^2_2$ is level 2 variance from unconditional model,

$\tau^2_3$ is level 3 variance from unconditional model,

$\rho_2 = \frac{\tau^2_2}{\tau^2_2 + \tau^2_3 + \sigma^2}$ and represents proportion of variance in the outcome between level 2 units, $\rho_3 = \frac{\tau^2_3}{\tau^2_2 + \tau^2_3 + \sigma^2}$ and represents proportion of variance in the outcome between level 3 units,

$\sigma|X|^2$ is level 1 variance conditional on level 3 variables,

$\tau^2_3|V|$ is level 3 variance conditional on level 3 variables,

$\tau^2_2|W|$ is level 2 variance conditional on level 2 variables,

$R^2_1 = 1 - \sigma^2|X|/\sigma^2$ and is level 1 variance explained by level 1 variables,

$D = \frac{1}{1 - \rho^2_{TZ}}$ represents efficiency loss in comparison to a randomized experiment with the same model but no correlation between the assignment variable and treatment status,

$\rho^2_{TZ} = \frac{\sigma_{TZ}}{\sqrt{\sigma^2_T(1-P^2)\sigma^2_Z}}$, $\sigma_{TZ}$ is covariance between $T$ and $Z$ and $\sigma_Z$ is standard deviation of $Z$.

The hypothesis test for the treatment effect is conducted using a non-central t-distribution with non-centrality parameter $\frac{\xi_{100}}{\sqrt{\text{var}(\xi_{100})}}$ with degrees of freedom $K(nJ - 1 - (g_1 + g_2) - 2$ where $g_1$ and $g_2$ are number of covariates included at level 2 and level 3 other than treatment and assignment variables.
Monte Carlo simulation results for BIRD3rc2rc1 design are presented in the Appendix A, Table A.1.1 and A.1.2. Results indicate that there is close correspondence between \( SE_{\text{emp}}(\xi_{100}) \) and \( SE_{\text{eq}}(\xi_{100}) \). Similarly, result indicate that there is close correspondence between empirical power and calculated power (See Table A.1.1). However, relative difference between empirical power and calculated power values tends to increase as top-level sample size become smaller. As for comparison of empirical and calculated Type I error rates, they are similar to comparison of empirical power and calculated power (See Table A.1.2). Unlike the relative difference in empirical power and calculated power values, the relative difference in the empirical Type I error and calculated Type I error rates does not systematically depend on top-level sample size. Finally, ninety-five percent CI formula coverage rates hover around the expected 95%.

### 4.1.2.1 Derivation

Different from design 4.1.1 vector of coefficients is

\[
\beta_{jk} = \begin{pmatrix}
\xi_{000} + \mu_{0jk} + \zeta_{00k} \\
\xi_{100} \\
\xi_{200}
\end{pmatrix}.
\]

Recall

\[
\text{Var}(\beta_{jk}) = E \left( \text{Var}(\beta_{jk} | \Omega) \right) + \text{Var} \left( E(\beta_{jk} | \Omega) \right). \tag{4.15}
\]

Focusing on the treatment effect
\[ \text{Var}(\hat{\xi}_{100}) = E\left(\text{Var}(\hat{\xi}_{100}|\omega_{22})\right) + \text{Var}\left(E(\hat{\xi}_{100}|\omega_{22})\right). \] (4.16)

The variance of the conditional OLS estimator is same as in design BIRD3rr2rr1, so Equation 4.16 can be stated as

\[ \text{Var}(\hat{\xi}_{100}) = \sigma(X_j^T X_{jk})^{-1} + \omega_{22}, \] (4.17)

where

\[ X_{jk} = \begin{bmatrix} 1_{n^P} & (1-P)1_{nP} & (Z - Z_0)_{nP} \\ 1_{n(1-P)} & -P1_{n(1-P)} & (Z - Z_0)_{n(1-P)} \end{bmatrix}. \]

However the variance that comes from design effect differs as OLS estimator of treatment effects for each sub-cluster assumed to be same so the structure of the \( \Omega \)

\[ \text{Var}\left(E(\hat{\xi}^{2}|\Omega)\right) = \Omega = \begin{bmatrix} \frac{\tau_{2W}^{2}}{JK} + \frac{\tau_{3V}^{2}}{K} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \]

Focusing on the treatment effect variance

\[ \text{Var}\left(E(\hat{\xi}_{100}|\omega_{22})\right) = \omega_{22} = 0. \]

Therefore
\[ \text{Var}(\hat{\xi}_{100}) = \frac{\sigma_{\hat{\xi}}^2}{nJKP(1-P)(1-p_{2T})}. \]  (4.18)

Considering that \( R_1^2 = 1 - \sigma_{\hat{\xi}}^2/\sigma^2 \), \( \rho_2 = \frac{\tau_2^2}{\tau_2^2 + \tau_2^3 + \sigma^2} \), \( \rho_3 = \frac{\tau_3^2}{\tau_3^2 + \tau_2^3 + \sigma^2} \) and by standardizing the equality by \( \tau_3^2 + \tau_2^2 + \sigma^2 \) the Equation 4.18 becomes

\[ \frac{\text{Var}(\hat{\xi}_{100})}{\tau_3^2 + \tau_2^2 + \sigma^2} = \frac{(1-P_2-P_3)(1-R_1^2)}{nJKP(1-P)(1-p_{2T})}. \]  (4.19)

### 4.1.3 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts Mean Effects are Random but Treatment Effect is Random and Level 3 Mean Effects are Fixed but Treatment Effect is Constant (BIRD3fc2rr1)

Researcher may want to design a three-level regression discontinuity studies where the treatment variable is at level 1, and where level 2 intercepts vary randomly but level 3 intercepts are fixed, and treatment effect is constant across level 3 units but random across level 2 units. In this case, the sampling variance of the treatment is proportional to (un)conditional residual variance at level 1 and level 2, and inversely proportional to sample sizes at each level, variance of the treatment variable, and the strength of the relationship between the treatment and assignment variables.

Since treatment effect is constant across level 2 and level 3 units, compared to BIRD3rr2rr1 model, the terms \( \omega_3 = 0 \) and \( R_3^2 = 1 \). The unstandardized form of the sampling variance of the treatment effect becomes
\[ Var(\hat{\tau}_{100}) = \frac{\tau^2_{2|W}}{JK} + \frac{D\sigma^2_{1X}}{nJKP(1-P)}, \]  

(4.20)

and in the standardized form

\[ \frac{Var(\hat{\tau}_{100})}{\tau^2 + \sigma^2} = \frac{\rho_2 \omega_2 (1 - R^2_{2T})}{JK} + \frac{D(1 - \rho_2)(1 - R^2_{1})}{nJKP(1-P)}, \]  

(4.21)

where

\[ \rho_2 = \frac{\tau^2_2}{\tau^2 + \sigma^2} \] and represents proportion of variance in the outcome explained by level 2 units,

\[ \omega_2 = \frac{\tau^2_2}{\tau^2} \] and represents treatment effect heterogeneity across level 2 units,

\( \sigma^2 \) is level 1 variance from unconditional model,

\( \tau^2_2 \) is level 2 variance from unconditional model,

\( \sigma_{1|X}^2 \) is level 1 variance conditional on level 3 variables,

\( \tau^2_{2|W} \) is level 2 variance conditional on level 2 variables,

\( R^2_1 = 1 - \sigma^2_{1|X}/\sigma^2 \) and is level 1 variance explained by level 1 variables,

\( R^2_{2T} = 1 - \tau^2_{2|W}/\tau^2_{2T} \) and is proportion of variance at level 2 on the treatment explained by level 2 variables,
\[ D = \frac{1}{1-\rho_{TZ}^2} \] represents efficiency loss in comparison to a randomized experiment with the same model but no correlation between the assignment variable and treatment status, \( \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{\text{Var}(T)\text{Var}(Z)}} \), \( \sigma_{TZ} \) is covariance between \( T \) and \( Z \) and \( \sigma_Z \) is standard deviation of \( Z \).

The hypothesis test for the treatment effect is conducted using a non-central t-distribution with non-centrality parameter \( \frac{\hat{\xi}_{100}}{\sqrt{\text{Var}(\hat{\xi}_{100})}} \) with degrees of freedom \( JK(n - 1) - (J - 2) - g_2 \) where \( g_2 \) is number of covariates included at level 2.

Monte Carlo simulation results for BIRD3fc2rr1 design are presented in the Appendix A, Table A.1.3 and A.1.4. Results indicate that there is close correspondence between \( \text{SE}_{\text{emp}}(\hat{\xi}_{100}) \) and \( \text{SE}_{\text{eq}}(\hat{\xi}_{100}) \), empirical power and calculated power, and empirical and calculated Type I error rates. The relative difference between the empirical and calculated power values tends to increase as top-level sample size become smaller whereas relative difference between the empirical and calculated Type I error does not seem to depend on top-level sample size systematically. Finally, ninety-five percent CI formula coverage rates hover around expected 95%.

### 4.1.3.1 Derivation

Different from design 4.1.1 vector of coefficients is

\[
\beta_{jk} = \begin{pmatrix}
\hat{\xi}_{000} + \mu_{0jk} \\
\hat{\xi}_{100} + \mu_{1jk} \\
\hat{\xi}_{200}
\end{pmatrix}.
\]
Recall

\[
\text{Var}(\hat{\beta}_{jk}) = E(\text{Var}(\hat{\beta}_{jk} | \Omega)) + \text{Var}(E(\hat{\beta}_{jk} | \Omega)). \tag{4.22}
\]

Focusing on the treatment effect

\[
\text{Var}(\hat{\xi}_{100}) = E(\text{Var}(\hat{\xi}_{100} | \omega_{22})) + \text{Var}(E(\hat{\xi}_{100} | \omega_{22})). \tag{4.23}
\]

The variance that comes from sampling of the subjects from the population is same as design 4.1.1, so this expression can be stated as

\[
\text{Var}(\hat{\xi}_{100}) = \sigma(X_{jk}^T X_{jk})^{-1} + \omega_{22}. \tag{4.24}
\]

where

\[
X_{jk} = \begin{bmatrix}
1_{nP} & (1 - P)1_{nP} & (Z - Z_0)_{nP} \\
1_{n(1-P)} & -P1_{nP(1-P)} & (Z - Z_0)_{nP(1-P)}
\end{bmatrix}
\]

However the variance that comes from design effect differs as OLS estimator of treatment effects for each sub-cluster assumed to be same so the structure of the \( \Omega \)

\[
\text{Var}\left(E(\hat{\xi} | \Omega)\right) = \Omega = \begin{bmatrix}
\frac{\tau^2_{2W}}{JK} & \frac{\tau_{2T2W}}{JK} & 0 \\
\frac{\tau_{2T2W}}{JK} & \frac{\tau^2_{2T1W}}{JK} & 0 \\
0 & 0 & 0
\end{bmatrix}
\]
Focusing on the treatment effect variance

\[ \text{Var} \left( E(\hat{\xi}_{100} | \omega_{22}) \right) = \omega_{22} = \frac{\tau^2_{T|W}}{JK}. \]

Therefore

\[ \text{Var}(\hat{\xi}_{100}) = \frac{\tau^2_{T|W}}{JK} + \frac{\sigma^2_{X}}{nJK(1-P)(1-\rho^2_{2T})}. \] (4.25)

Recalling that \( R_1^2 = 1 - \sigma^2_{|X}/\sigma^2 \), and \( R^2_{2T} = 1 - \tau^2_{T|W}/\tau^2_{T} \) and \( \rho_2 = \frac{\tau^2}{\tau^2 + \sigma^2} \) by standardizing the equality by \( \tau^2 + \sigma^2 \) the equality becomes

\[ \frac{\text{Var}(\hat{\xi}_{100})}{\tau^2 + \sigma^2} = \frac{\rho_2 \omega_2 (1-R^2_{2T})}{JK} + \frac{(1-\rho_2-\rho_3)(1-R^2_1)}{nJK(1-P)(1-\rho^2_{2T})}. \] (4.26)
4.1.4 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts Mean Effects are Random but Treatment Effect Constant and Level 3 Mean Effects are Fixed but Treatment Effect is Constant (BIRD3fc2rc1)

Researcher may want to design a three-level regression discontinuity studies where the assignment variable is at level 1, and where level 2 intercepts vary randomly but level 3 intercepts are fixed, and treatment effect is constant across level 2 and level 3 units. In this case, the sampling variance of the treatment is proportional to (un)conditional residual variance, and inversely proportional to sample sizes at each level and variance of the treatment variable and the strength of the relationship between the treatment and assignment variables.

Since treatment effect is constant across level 2 and level 3 units, compared to BIRD3rr2rr1 model, the terms $\omega_2 = 0$, $\omega_3 = 0$, $R_{2T}^2 = 1$, and $R_{3T}^2 = 1$, furthermore since level 3 means are fixed $\rho_3 = 0$. Then the unstandardized form of the sampling variance of the treatment effect becomes

$$
Var(\hat{\epsilon}_{100}) = \frac{D\sigma^2_X}{nJKP(1-P)},
$$

(4.27)

and in the standardized form
\[
\frac{\text{Var}(\hat{\xi}_{100})}{\tau^2_2 + \sigma^2} = \frac{D(1-\rho_2)(1-R^2_1)}{nJKP(1-P)},
\]

where

\[
\rho_2 = \frac{\tau^2_2}{\tau^2_2 + \tau^2_3 + \sigma^2}
\]

and represents proportion of variance in the outcome explained by level 2 units,

\[
\sigma^2
\]

is level 1 variance from unconditional model,

\[
\tau^2_2
\]

is level 2 variance from unconditional model,

\[
\sigma_{\mid X}^2
\]

is level 1 variance conditional on level 3 variables,

\[
\tau^2_2\mid W
\]

is level 2 variance conditional on level 2 variables,

\[
R^2_1 = 1 - \sigma_{\mid X}^2/\sigma^2
\]

and is level 1 variance explained by level 1 variables,

\[
D = \frac{1}{1-P^2_\tau Z}
\]

represents efficiency loss in comparison to a randomized experiment with the same model but no correlation between the assignment variable and treatment status,

\[
\rho^2_{\tau Z} = \frac{\sigma_{\tau Z}}{\sqrt{P(1-P)\sigma^2_Z}}, \quad \sigma_{\tau Z} \text{ is covariance between } T \text{ and } Z \text{ and } \sigma^2_Z \text{ is standard deviation of } Z.
\]

The hypothesis test for the treatment effect is conducted using a non-central t-
distributuion with non-centrality parameter \(\frac{\hat{\xi}_{100}}{\sqrt{\text{Var}(\hat{\xi}_{100})}}\) with degrees of freedom \(JK(n - 1) - g_2 - 2\).

Monte Carlo simulation results for BIRD3fc2rc1 design are presented in the Appendix A, Table A.1.5 and A.1.6. Results indicate that there is close correspondence
between $SE_{emp}(\hat{\xi}_{100})$ and $SE_{eq}(\hat{\xi}_{100})$, empirical power and calculated power, and empirical and calculated Type I error rates. As with the other designs, the relative difference between the empirical and calculated power values tends to increase as top-level sample size become smaller whereas relative difference between the empirical and calculated Type I error does not seem to depend on top-level sample size systematically. Finally, ninety-five percent CI formula coverage rates hover around the expected 95%.

### 4.1.4.1 Derivation

Different from design 4.1.1 vector of coefficients is

$$\beta_{jk} = \begin{pmatrix} \hat{\xi}_{000} + \mu_{0jk} \\ \hat{\xi}_{100} \\ \hat{\xi}_{200} \end{pmatrix}.$$  

Recall

$$Var(\hat{\beta}_{jk}) = E \left( Var(\hat{\beta}_{jk}|\Omega) \right) + Var \left( E(\hat{\beta}_{jk}|\Omega) \right).$$  

(4.29)

Focusing on the treatment effect

$$Var(\hat{\xi}_{100}) = E \left( Var(\hat{\xi}_{100}|\omega_{22}) \right) + Var \left( E(\hat{\xi}_{100}|\omega_{22}) \right).$$  

(4.30)

The variance that comes from sampling of the subjects from the population is same as design 4.1.1, so this expression can be stated as
\[
V ar(\hat{\theta}_{100}) = \sigma (X_{jk}^T X_{jk})^{-1} + \omega_{22}. \tag{4.31}
\]

where

\[
X_{jk} = \begin{bmatrix}
1_{nP} & (1-P)1_{nP} & (Z - Z_0)_{nP} \\
1_{n(1-P)} & -P1_{n(1-P)} & (Z - Z_0)_{n(1-P)}
\end{bmatrix}.
\]

However the variance that comes from design effect differs as OLS estimator of treatment effects for each sub-cluster assumed to be same so the structure of the \( \Omega \)

\[
V ar\left( E(\hat{\theta} | \Omega) \right) = \Omega = \begin{bmatrix}
\tau^2_{i1w} & 0 & 0 \\
0 & jk & 0 \\
0 & 0 & 0
\end{bmatrix}.
\]

Focusing on the treatment effect variance

\[
V ar\left( E(\hat{\theta}_{100} | \omega_{22}) \right) = \omega_{22} = 0.
\]

Therefore

\[
V ar(\hat{\theta}_{100}) = \frac{\sigma^2}{nJKP(1-P)(1-\rho^2_2)} \tag{4.32}
\]

Recalling that \( R_1^2 = 1 - \sigma^2_1 / \sigma^2 \), and \( \rho_2 = \frac{\tau^2_e}{\tau^2_e + \sigma^2} \) and by standardizing the equality by \( \tau^2_2 + \sigma^2 \) the equality becomes
\[
\frac{\text{Var}(\hat{\xi}_{100})}{\tau_2^2 + \sigma^2} = \frac{(1-\rho_2)(1-R_1^2)}{nJK(1-P)(1-\rho_T^2)}.
\] (4.33)
4.1.5 Three-level RD Study where Level 1 Units are Unit of Assignment, Level 2 Intercepts and Level 3 Mean Effects are Fixed and Treatment Effect is Constant (BIRD3fc2fc1)

Researcher may want to design a three-level regression discontinuity study where the assignment variable is at level 1, and where level 2 and level 3 intercepts are fixed, and treatment effect is constant across level 2 and level 3 units. In this case, the sampling variance of the treatment is proportional to (un)conditional residual variance, and inversely proportional to sample sizes at each level and variance of the treatment variable and the strength of the relationship between the treatment and assignment variables.

Since treatment effect is constant across level 2 and level 3 units, compared to BIRD3rr2rr1 model, the terms $\omega_2 = 0$, $\omega_3 = 0$, $R_{2T}^2 = 1$, and $R_{3T}^2 = 1$, furthermore since level 2 and level 3 means are fixed $\rho_2 = 0$ and $\rho_3 = 0$. Then the unstandardized form of the sampling variance of the treatment effect become

$$Var(\bar{\xi}_{100}) = \frac{D\sigma_{|X|^2}}{nJKP(1-P)}, \quad (4.34)$$

and in the standardized form

$$\frac{Var(\bar{\xi}_{100})}{\tau_2^2 + \sigma^2} = \frac{D(1-R_{1T}^2)}{nJKP(1-P)}, \quad (4.35)$$

where

$\sigma^2$ is level 1 variance from unconditional model,

$\tau_2^2$ is level 2 variance from unconditional model,

$\sigma_{|X|^2}$ is level 1 variance conditional on level 3 variables,
\[ R_1^2 = 1 - \frac{\sigma_{ix}^2}{\sigma^2} \] and is level 1 variance explained by level 1 variables,

\[ D = \frac{1}{1 - \rho_{TZ}^2} \] represents efficiency loss in comparison to a randomized experiment with the same model but no correlation between the assignment variable and treatment status,

\[ \rho_{TZ}^2 = \frac{\sigma_{TZ}}{\sqrt{p(1-p)\sigma_z}} \] is covariance between \( T \) and \( Z \) and \( \sigma_z \) is standard deviation of \( Z \).

The hypothesis test for the treatment effect is conducted using a non-central \( t \)-distribution with non-centrality parameter \( \frac{\xi_{100}}{\sqrt{Var(\xi_{100})}} \) with degrees of freedom \( JK(n - 1) - g_1 - 2 \).

Monte Carlo simulation results for BIRD3fc2fc1 design are presented in the Appendix A, Table A.1.7 and A.1.8. Results indicate that there is close correspondence between \( SE_{emp}(\xi_{100}) \) and \( SE_{eq}(\xi_{100}) \), empirical power and calculated power values, empirical Type I error and calculated Type I error rates. As with the other designs, the relative difference between empirical and calculated power values tends to be big when top-level sample size is small. Unlike relative difference between empirical power and calculated power values, relative difference in empirical Type I error and calculated Type I error rates does not change systematically due to top-level sample size. Finally, ninety-five percent CI formula coverage rates hover around the expected 95%.

### 4.1.5.1 Derivation

Different from design 4.1.1 vector of coefficients is
\[ \mathbf{\beta}_{jk} = \begin{pmatrix} \xi_{000} \\ \xi_{100} \\ \xi_{200} \end{pmatrix}. \]

Recall

\[ \text{Var}(\hat{\mathbf{\beta}}_{jk}) = E\left(\text{Var}(\hat{\mathbf{\beta}}_{jk}|\mathbf{\Omega})\right) + \text{Var}(E(\hat{\mathbf{\beta}}_{jk}|\mathbf{\Omega})). \]

(4.36)

Focusing on the treatment effect

\[ \text{Var}(\hat{\xi}_{100}) = E\left(\text{Var}(\hat{\xi}_{100}|\omega_{22})\right) + \text{Var}(E(\hat{\xi}_{100}|\omega_{22})). \]

(4.37)

The variance that comes from sampling of the subjects from the population is same as design 4.1.1, so this expression can be stated as

\[ \text{Var}(\hat{\xi}_{100}) = \sigma(\mathbf{X}_j\mathbf{kX}_j)^{-1} + \omega_{22}. \]

(4.38)

where

\[
\mathbf{X}_{jk} = \begin{bmatrix}
\mathbf{1}_{nP} \\
\mathbf{1}_{n(1-P)}
\end{bmatrix} \begin{bmatrix}
(1 - P)\mathbf{1}_{nP} \\
-Z\mathbf{0}_{nP} \\
-Z\mathbf{0}_{n(1-P)}
\end{bmatrix}.
\]

However the variance that comes from design effect differs as OLS estimator of treatment effects for each sub-cluster assumed to be same so the structure of the \( \mathbf{\Omega} \)

\[ \text{Var}(E(\hat{\xi}|\mathbf{\Omega})) = \mathbf{\Omega} = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}. \]
Focusing on the treatment effect

\[ \text{Var} \left( E(\hat{\xi}_{100} | \omega_{22}) \right) = \omega_{22} = 0. \]

Therefore

\[ \text{Var}(\hat{\xi}_{100}) = \frac{\sigma_{\hat{\xi}}^2}{nJKP(1-P)(1-\rho_{Z\tau}^2)}. \] (4.39)

Recalling that \( R_1^2 = 1 - \frac{\sigma_X^2}{\sigma^2} \), and by standardizing the equality by \( \sigma^2 \) the equality becomes

\[ \frac{\text{Var}(\hat{\xi}_{100})}{\sigma^2} = \frac{(1-R_1^2)}{nJKP(1-P)(1-\rho_{Z\tau}^2)}. \] (4.40)
4.2 Omission of Intermediate Level in BIRD3rr2rr1 Model

When the intermediate level is ignored in a BIRD3rr2rr1 model, it becomes a two-level model (BIRD2rr1) where the previous third level remains as top level. In addition to the shift in the variance components which affects the new top and bottom level, sample size for top level remains the same ($K$), but sample size for level 1 is now the combined sample size ($nJ$) whereas degrees of freedom for the test statistics does not change. In this case, power is slightly underestimated. On the other hand, when the top level is ignored, in addition to the shift in variance components which affects the level below the ignored level, sample size for the new top level is now combined ($JK$), however, sample size for the new bottom level remains the same ($n$) whereas degrees of freedom for the test statistics change due to change in number of top levels. As top-level sample size is one of the most important determinant of power, the change in top-level sample size alone is sufficient to overestimate power. Monte Carlo simulation result confirms these findings and are presented in Tables 4.3 and 4.4. Results in Tables 4.3 and 4.4 should be interpreted along with the results in Tables 4.1 and 4.2.
Table 4.3 Power Analysis for the Misspecified BIRD2rr1 Model (Level 2 Ignored)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\xi}_{100}$</td>
<td>0.24</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$SE(\hat{\xi}_{100})$</td>
<td>0.15</td>
<td>0.18</td>
<td>0.11</td>
<td>0.13</td>
<td>0.15</td>
<td>0.20</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>ES(\hat{\xi}_{100})</td>
<td>0.10</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
<td>0.12</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.30</td>
<td>0.12</td>
<td>0.30</td>
<td>0.12</td>
<td>0.30</td>
<td>0.11</td>
<td>0.30</td>
<td>0.11</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.54</td>
<td>0.49</td>
<td>0.53</td>
<td>0.48</td>
<td>0.66</td>
<td>0.57</td>
<td>0.65</td>
<td>0.57</td>
</tr>
<tr>
<td>$R^2_1$</td>
<td>0.22</td>
<td>0.04</td>
<td>0.22</td>
<td>0.04</td>
<td>0.22</td>
<td>0.03</td>
<td>0.22</td>
<td>0.03</td>
</tr>
<tr>
<td>$R^2_{2T}$</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>$\rho_T$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>$K$</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

| $SE_{emp}(\hat{\xi}_{100})$ | 0.15 | 0.19 | 0.11 | 0.13 | 0.15 | 0.21 | 0.11 | 0.14 |
| $SE_{eq}(\hat{\xi}_{100})$ | 0.15 | 0.19 | 0.11 | 0.14 | 0.15 | 0.20 | 0.11 | 0.14 |
| AD in SEs | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| RD in SEs (%) | -1.26 | 2.86 | -0.95 | 5.06 | -0.31 | -2.26 | -1.62 | -0.40 |
| 95% CI Coverage | 0.95 | 0.96 | 0.95 | 0.96 | 0.95 | 0.95 | 0.94 | 0.95 |
| Empirical Power | 0.38 | 0.28 | 0.65 | 0.49 | 0.38 | 0.24 | 0.62 | 0.42 |
| Calculated Power | 0.35 | 0.24 | 0.64 | 0.44 | 0.36 | 0.22 | 0.61 | 0.41 |
| AD in Powers | -0.03 | -0.04 | -0.01 | -0.05 | -0.02 | -0.02 | -0.01 | -0.01 |
| RD in Powers (%) | -6.78 | -13.68 | -2.24 | -10.35 | -4.55 | -8.56 | -1.51 | -1.25 |

Note. Results are based on 5000 replications. $\hat{\xi}_{100}$: Treatment effect. SE: Standard Error. ES: Effect size. $\rho_2$: Proportion of variance in the outcome between level 2 units. $\omega_2$: Treatment effect heterogeneity across level 2 units. $R^2_1$: Proportion of variance in the outcome explained level 1 covariates. $R^2_{2T}$: Proportion of variance in the treatment effect explained level 2 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $\rho_T$: Correlation between the assignment variable and the treatment status. $nJ$: Average number of level 1 units per level 2 units, which is set to 100. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
Table 4.4 Power Analysis for the Misspecified BIRD2rr1 Model (Level 3 Ignored)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\xi}_{100}$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>SE ($\hat{\xi}_{100}$)</td>
<td>0.10</td>
<td>0.14</td>
<td>0.07</td>
<td>0.10</td>
<td>0.11</td>
<td>0.17</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>ES ($\hat{\xi}_{100}$)</td>
<td>0.10</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
<td>0.12</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.61</td>
<td>0.24</td>
<td>0.61</td>
<td>0.24</td>
<td>0.59</td>
<td>0.22</td>
<td>0.59</td>
<td>0.22</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.65</td>
<td>0.52</td>
<td>0.65</td>
<td>0.52</td>
<td>0.77</td>
<td>0.60</td>
<td>0.77</td>
<td>0.59</td>
</tr>
<tr>
<td>$R_1^2$</td>
<td>0.53</td>
<td>0.07</td>
<td>0.54</td>
<td>0.07</td>
<td>0.48</td>
<td>0.05</td>
<td>0.48</td>
<td>0.05</td>
</tr>
<tr>
<td>$R_2^2_{tr}$</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$\rho_{TZ}$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>JK</td>
<td>250</td>
<td>250</td>
<td>500</td>
<td>500</td>
<td>250</td>
<td>250</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

SE$_{emp}$ ($\hat{\xi}_{100}$) | 0.14 | 0.19 | 0.10 | 0.13 | 0.14 | 0.20 | 0.10 | 0.14 |
SE$_{eq}$ ($\hat{\xi}_{100}$) | 0.10 | 0.16 | 0.07 | 0.11 | 0.10 | 0.17 | 0.07 | 0.12 |
AD in SEs | -0.04 | -0.03 | -0.03 | -0.02 | -0.04 | -0.03 | -0.03 | -0.02 |

95% CI Coverage | 0.85 | 0.90 | 0.84 | 0.90 | 0.85 | 0.90 | 0.84 | 0.90 |
Empirical Power | 0.62 | 0.43 | 0.86 | 0.66 | 0.63 | 0.34 | 0.84 | 0.54 |
Calculated Power | 0.66 | 0.33 | 0.93 | 0.60 | 0.68 | 0.31 | 0.92 | 0.56 |
AD in Powers | 0.04 | -0.10 | 0.07 | -0.05 | 0.04 | -0.03 | 0.08 | 0.01 |
RD in Powers (%) | 6.16 | -22.61 | 8.72 | -8.28 | 6.79 | -8.79 | 10.12 | 2.41 |

Note. Results are based on 5000 replications. $\hat{\xi}_{100}$: Treatment effect. SE: Standard Error. ES: Effect size. $\rho_2$: Proportion of variance in the outcome between level 2 units. $\omega_2$: Treatment effect heterogeneity across level 2 units. $R_1^2$: Proportion of variance in the outcome explained level 1 covariates. $R_2^2_{tr}$: Proportion of variance in the treatment effect explained level 2 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $\rho_{TZ}$: Correlation between the assignment variable and the treatment status. n: Average number of level 1 units per level 2 units, which is set to 20. JK: Number of level 2 units. AD: Absolute difference. RD: Relative difference.
Table 4.5 demonstrates how variance parameters for unconditional model shifts when intermediate- or top-level is ignored using the Monte Carlo simulation. When the intermediate level is ignored in the BRID3rr2rr1 model, the level 2 variance is distributed to the flanking levels in the new BIRD2rr1 model. The variance distributed to the bottom level model is greater than the variance distributed to the top level in the new BIRD2rr1 model. When the top level is ignored, the bottom level remains the same, however, level 2 variance in the new BIRD2rr1 model is inflated approximately equal to the sum of level 2 and level 3 variance in the BIRD3rr2rr1 model.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>True model</td>
<td>$\sigma^2$</td>
<td>2.15</td>
<td>9.66</td>
<td>2.15</td>
<td>9.66</td>
<td>1.92</td>
<td>9.49</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>$\tau_2^2$</td>
<td>2.08</td>
<td>1.89</td>
<td>2.07</td>
<td>1.90</td>
<td>1.69</td>
<td>1.64</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>$\tau_3^2$</td>
<td>1.27</td>
<td>1.21</td>
<td>1.27</td>
<td>1.21</td>
<td>1.11</td>
<td>1.08</td>
<td>1.11</td>
</tr>
<tr>
<td>Level 2 ignored</td>
<td>$\sigma^2$</td>
<td>3.83</td>
<td>11.18</td>
<td>3.83</td>
<td>11.19</td>
<td>3.29</td>
<td>10.81</td>
<td>3.29</td>
</tr>
<tr>
<td>misspecified model</td>
<td>$\tau_2^2$</td>
<td>1.66</td>
<td>1.58</td>
<td>1.67</td>
<td>1.58</td>
<td>1.44</td>
<td>1.39</td>
<td>1.43</td>
</tr>
<tr>
<td>Level 3 ignored</td>
<td>$\sigma^2$</td>
<td>2.15</td>
<td>9.66</td>
<td>2.15</td>
<td>9.66</td>
<td>1.92</td>
<td>9.49</td>
<td>1.92</td>
</tr>
<tr>
<td>misspecified model</td>
<td>$\tau_2^2$</td>
<td>3.32</td>
<td>3.08</td>
<td>3.33</td>
<td>3.10</td>
<td>2.78</td>
<td>2.69</td>
<td>2.79</td>
</tr>
</tbody>
</table>

*Note.* Same symbols bear different meaning in different models. $\sigma^2$: Level 1 variance. $\tau_2^2$: Level 2 variance. $\tau_3^2$: Level 3 variance.
4.3 Constrained Optimal Sample Allocation

4.3.1 Introduction to PowerUpR R Package

PowerUpR is an R package that provides tools to solve constrained optimal sample allocation (COSA) problems for both randomized experiments (RE) and regression discontinuity (RD) designs. COSA problems can be solved (i) under budgetary constraints given marginal costs per treatment and control units while minimizing sampling variance of the treatment effect, (ii) under power or MDES constraints given marginal costs per treatment and control units while minimizing the total cost, and (iii) under sample size constraints for one or more levels along with any of the i, ii, or iii options.

COSA functions begin with the cosa keyword, following by a period, and a design name. Along with various multilevel RE designs there are eight multilevel RD designs included in PowerUpR package: bird2r1, bird2f1, crd2r2, bird3r1, bird3c1, bird3r2, bird3f2, and crd3r3. The first three or four letters of the design stands for the type of assignment, for block/multisite individual-level regression discontinuity bird, for cluster-level regression discontinuity crd, for block/multisite cluster-level regression discontinuity bcrd. The number that follows indicates number of levels. The single letter between the two numbers indicates whether the intercept and treatment slope is; r for random, f for fixed, or c for constant across higher level units. The last number indicates the level at which the discontinuity occurs. For example, to solve COSA problem for a three-level cluster-level regression discontinuity design, the cosa.crd3r3 function is used.
Each COSA function requires slightly different arguments depending on the design. Most of the arguments have default values to provide users a starting point.

**Default values are:** `mdes= 0.25, power=0.80, alpha=0.05, two.tail=TRUE, p=0.50, n=NULL, J=NULL, K=NULL, nJK0=c(10, 10, 10), RTZ= NULL, k1=-6, k2=6, dist.Z="normal", optimizer="auglag_slsqp", g1=0, g2=0, g3=0, R12=0, R22=0, R32=0, RT22=0, and RT32=0. Users should be aware of default values and change them if necessary. Minimum required arguments to successfully run a function are: any sequence of `rho2`, `rho3`, `omega2`, `omega3`, and any one of, any sequence of, or any combination of `n`, `J`, `K`. For definition of above-mentioned parameters see Dong & Maynard (2013) and Hedges & Rhoads (2009), or help files in the PowerUpR package. For reference intraclass correlation (`rho2`, `rho3`) values see Deke, Dragoset, and Moore (2010), Dong, Reinke, Herman, Bradshaw, and Murray (2016), Hedberg and Hedges (2014), Hedges and Hedberg (n.d., 2007, 2013), Kelcey, and Phelps (2013), Schochet (2008a), Spybrook, Westine, and Taylor (2016). For reference variance (`R12`, `R22`, `R32`) values see Bloom, Richburg-Hayes, and Black (2007), Deke et al. (2010), Dong et al. (2016), Hedges and Hedberg (2013), Kelcey, and Phelps (2013), Spybrook, Westine, and Taylor (2016), Westine, Spybrook, and Taylor (2013). Users can also obtain design parameters for various levels using publicly available state or district data.

### 4.3.2 Example

Upon unsatisfactory 1st grade reading scores in a recent national test, assume a state education agency is planning to allocate funding to underperforming public
elementary schools to prioritize reading activities with during and after school programs focusing on 1st graders. Furthermore, assume schools that have performed below national average will be granted funding. The state education agency would like to evaluate the impact of such programs before deploying funding at the state level.

4.3.2.1 Statistical Model

Due to nesting of students within classrooms and nesting of classrooms within schools, the treatment effect (\( \delta \)) can be estimated via the following three-level hierarchical linear model (Raudenbush & Bryk, 2002)

- **Level 1:** \( Y_{ijk} = \beta_{0jk} + \beta_{1jk}X_{ijk} + r_{ijk} \)
- **Level 2:** \( \beta_{0jk} = \gamma_{00k} + \gamma_{01k}W_{jk} + \mu_{0jk} \)

  \[ \beta_{1jk} = \gamma_{10k} \]

- **Level 3:** \( \gamma_{00k} = \xi_{000} + \delta T_k + \xi_{002}(Z_k - Z_0) + \zeta_{00k} \)

  \[ \gamma_{01k} = \xi_{010} \]

  \[ \gamma_{10k} = \xi_{100} \]

where \( r_{ijk} \sim N(0, \sigma_r^2) \), \( \mu_{0jk} \sim N(0, \tau^2_{\mu}) \), \( \xi_{000} \sim N(0, \tau^2_{\xi}) \), \( \gamma_{00k} \sim N(0, \tau^2_\gamma) \), and \( Y_{ijk} \): Student reading posttest score. \( X_{ijk} \): Student reading pretest score. \( W_{jk} \): Classroom mean reading pretest score. \( T_k \): Treatment status derived from assignment variable \( Z_k \) based on the national average reading pretest score \( Z_0 \). Consider the following hypothetical values for design parameters

- Proportion of variance in reading posttest at the school level (\( \tau^2_3 \)): 0.15
• Proportion of variance in reading posttest at the classroom level ($\tau_2^2$): 0.05
• Proportion of variance in reading posttest at the student level ($\sigma^2$): 0.65
• Proportion of variance in reading posttest explained by school mean reading pretest ($R_3^2$): 0.45
• Proportion of variance in reading posttest explained by classroom mean reading pretest ($R_2^2$): 0.50
• Proportion of variance in reading posttest explained by student reading pretest ($R_1^2$): 0.55
• Proportion of schools that receives funding ($\mathbb{P}$): 0.50 (Assuming school mean reading pretest score follows a normal distribution and on average 50% of the schools falls below national average).

Then, intraclass correlation coefficients are
• $\rho_2 = \frac{\tau_2^2}{\sigma^2 + \tau_2^2 + \tau_3^2} = 0.059$
• $\rho_3 = \frac{\tau_3^2}{\sigma^2 + \tau_2^2 + \tau_3^2} = 0.176$

Further assumptions
• Alpha level ($\alpha$): 0.05
• Statistical power ($1 - \beta$): .80
• Minimum detectable effect size (MDES): .20
• The range for average number of students per classroom ($n$): 10 to 40
• The range for average number of classrooms per school ($J$): 2 to 4
• Marginal costs per student in the treatment and control conditions ($cn$): 15 and 10
• Marginal costs per classroom in the treatment and control conditions ($cJ$): 200 and 100
• Marginal cost per school in the treatment and control conditions (cK): 500 and 250

• Budget or total cost: 575,000

Given design parameters, further assumptions, budget, and marginal costs per units in treatment and control conditions, what is the optimal allocation of subjects/clusters across levels to achieve highest level of precision?

**Step 1**

```r
> library(PowerUpR)
> cosa.crd3r3(cn=c(15,10), cj=c(200,100), cK=c(500,250),
+ constrain="cost", cost=575000, n=c(10,40), j=c(2,4),
+ power=0.80, mdes=0.20, rho2=0.06, rho3=0.18,
+ g3=0, R12=0.55, R22=0.50, R32=0.45, P=0.50)
## Rounded solution:
## ---------------------------------------
##     n     J   K   P   cost  mdes  power
## 27.645 3.963 246 0.5 575425 0.197 0.812
## ---------------------------------------
## MDES is calculated with a power of 80% and
## power is calculated for an MDES of 0.2
```

**Step 2**

```r
> cosa.crd3r3(cn=c(15,10), cj=c(200,100), cK=c(500,250),
+ constrain="cost", cost=575000, n=28, j=4,
+ power=0.80, mdes=0.20, rho2=0.06, rho3=0.18,
+ g3=0, R12=0.55, R22=0.50, R32=0.45, P=.50)
## Rounded solution:
## ---------------------------------------
##    n J  K  P  cost  mdes  power
## 28 4 242 0.5 574750 0.199 0.806
## ---------------------------------------
## MDES is calculated with a power of 80% and
## power is calculated for an MDES of 0.2
```

Result in Step 2 indicate that with $575,000 an MDES of 0.20 can be detected with 80% statistical power. Optimal number of students, classrooms and schools that satisfy these conditions are 28, 4, and 242 respectively. In Step 1 n and J parameters are not integers because the formula assumes values entered for these parameters are averages or harmonic means. Users can fix n and J at their rounded values and obtain another
solution to avoid confusion. For the sake of comparison, how the result in Step 1 would have been different if schools were assigned to treatment and control conditions at random as opposed to based on national average? The following section addresses this comparison.

### 4.3.2.2 Comparison to Random Assignment

As the relationship between the treatment and assignment variables become weaker results approach to that of random assignment (RA) case. For identical results, the correlation between treatment status and assignment variable \((RTZ)\) is set to zero. Therefore, the following result is identical to RA case.

```r
> cosa.crd3r3(cn=c(15,10), cJ=c(200,100), cK=c(500,250), cost=575000, + constrain="mdes", mdes=0.197, n=28, J=4, + power=0.80, rho2=0.06, rho3=0.18, + g3=0, R12=0.55, R22=0.50, R32=0.45, P=0.50, RTZ=0)
```

```
## Rounded solution:
## ---------------------------------------
## n  J  K     P   cost  mdes  power
## 28 4 91 0.505 216730 0.197 0.802
## ---------------------------------------
## MDES is calculated with a power of 80% and
## power is calculated for an MDES of 0.197
```

Alternatively, researcher can use a function specific to RA case.

```r
> cosa.cra3r3(cn=c(15,10), cJ=c(200,100), cK=c(500,250), cost=575000, + constrain="mdes", mdes=0.197, n=28, J=4, + power=0.80, rho2=0.06, rho3=0.18, + g3=0, R12=0.55, R22=0.50, R32=0.45, P=0.50)
```

```
## Rounded solution:
## ---------------------------------------
## n  J  K     P   cost  mdes  power
## 28 4 91 0.505 216730 0.197 0.802
## ---------------------------------------
## MDES is calculated with a power of 80% and
## power is calculated for an MDES of 0.197
```
Results indicate that, if RA is feasible, to detect an MDES of 0.197 with 80% statistical power researcher needs to recruit 91 schools, which largely reduces the cost ($216,730).

4.3.2.3 Summarizing Results

A three-level hierarchical linear model is proposed to estimate the effect of state funding, where schools are assigned to treatment and control conditions based on their national average 1\textsuperscript{st} grade reading score. We assume that proportion of variance in reading posttest explained by classroom mean reading pretest is 50\%, proportion of variance in reading posttest explained by student reading pretest is 55\%, and proportion of variance in reading posttest explained by school reading pretest is 45\%. Furthermore, we assume intraclass correlation coefficient at the school level is 17.6\%, and intraclass correlation coefficient at the classroom level is 5.9\%.

Having $15 and $10 marginal cost per student in treatment and control conditions, $200 and $100 marginal cost per classroom in treatment and control conditions, $500 and $250 marginal cost per school in treatment and control conditions, with a budget of $575,000 we can afford recruiting 242 schools in total, each with four classrooms and 112 students on average. For a two-tailed hypothesis test for treatment effect, and Type I error rate of 5\%, such sample is estimated to have statistical power of 80\% to detect an effect size as small as .20.
5 Conclusion

This study derived formulas for three-level regression discontinuity (RD) studies where discontinuity resides at level 1 assuming that the generalized least square estimator of the treatment effect can be stated as an ordinary least square estimator conditional on level 2 and level 3 units. The average of these estimators are the average treatment effect and the variability in the slope of the coefficient contributes to the sampling variance of the average treatment effect. Across the five designs proposed in this study, the sampling variance component that emerges from the sampling of level 1 units (that from OLS estimation) remains the same. When higher level units are included in the model as fixed effects, the value of the intra class correlation or the variance explained by covariates are different from the model where higher level units are sampled randomly. Although the form of the first component in the formula remains the same, these variance components and the degrees of freedom alters the value of statistical power. The second component is the variance that emerge from variation in the slope coefficient and it varies from one design to another via imposing constraints on the variance components in the formulas. If the slope coefficient is considered to be constant or to vary non-randomly across higher-level units this component is zero.

It is apparent that one can use the most general formula from the model where the intercept and slope is random across level 2 and level 3 units (BIRD3rr2rr1) by constraining various variance parameters to either one or zero. This on its own is not sufficient because the degrees of freedom for the test statistics should also be altered due to fixed effects included in the model. Considering the two factors, the change in the variance parameters and in the degrees of freedom, one can design the remaining studies
based on the general formula. However, all designs are provided merely for convenience for the fact that it can be confusing to some researcher which parameters to constrain and how to modify the degrees of freedom. Monte Carlo simulation results indicates close correspondence between empirical and calculated standard errors, empirical and calculated power values, and empirical and calculated Type I error rates. Furthermore, 95% CI for formula coverage is hovering around the 95% in all cases. Degrees of freedom values for each design is confirmed using the SAS output from the simulation.

Statistical power values in Table 5.1 pertain to the most general design where intercept and slope for the treatment varies randomly across level 2 and level 3 units. The table compares statistical power for regression discontinuity (RD) and random assignment (RA) studies by specifying the correlation between the treatment and assignment variables to zero. In the table, the proportion of variance in the outcome that is explained by level 1 covariates $R^2_1$ varies whereas proportion of variance in the treatment slope that is explained by level 2 covariates $R^2_{2T}$, and the proportion of variance in the treatment slope that is explained by level 3 covariates $R^2_{3T}$ are constrained to be zero. Table 5.1 demonstrates that under similar assumptions an RA study have more statistical power in all conditions. The difference is more pronounced where the treatment split is not balanced ($P = 0.33$, and $P = 0.10$).

Another conclusion from the table is that exploratory power of covariates included at level 1 improves statistical power in RD more than it does in RA studies. To make up for the statistical power loss resulting from the correlation between the treatment and assignment variables in RD designs exploratory power of the covariate(s) should be improved significantly. Some parameters are related to correlation between treatment and
assignment variables in a direct manner. Therefore, changes in these parameters, if favors a weak correlation, partially diminishes discrepancy in the statistical power values from RA and RD designs. An assignment variable that is distributed uniformly and a treatment variable derived based on the $P = 0.10$ split indicates evidence in this direction.
Table 5.1 *Statistical Power for BIRD3rr2rr1 Design*

<table>
<thead>
<tr>
<th>Distribution of Z</th>
<th>$P = 0.50$</th>
<th>$P = 0.33$</th>
<th>$P = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_1$</td>
<td>RD</td>
<td>RA</td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.634</td>
<td>0.836</td>
<td>0.631</td>
</tr>
<tr>
<td>0.2</td>
<td>0.661</td>
<td>0.849</td>
<td>0.658</td>
</tr>
<tr>
<td>0.5</td>
<td>0.787</td>
<td>0.900</td>
<td>0.785</td>
</tr>
<tr>
<td>0.7</td>
<td>0.858</td>
<td>0.924</td>
<td>0.857</td>
</tr>
<tr>
<td>Uniform</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.531</td>
<td>0.836</td>
<td>0.631</td>
</tr>
<tr>
<td>0.2</td>
<td>0.561</td>
<td>0.849</td>
<td>0.609</td>
</tr>
<tr>
<td>0.5</td>
<td>0.712</td>
<td>0.900</td>
<td>0.750</td>
</tr>
<tr>
<td>0.7</td>
<td>0.810</td>
<td>0.924</td>
<td>0.835</td>
</tr>
</tbody>
</table>

*Note.* Z: Assignment variable. $P$: Proportion of level 1 units in treatment condition. $R^2_1$: Proportion of level 1 variance explained by covariates. RD: Regression Discontinuity. RA: Random Assignment. Values are based on an MDES = 0.20, proportion of variance in the outcome that is explained by level 2 clusters $\rho_2 = 0.20$, proportion of variance in the outcome that is explained by level 3 clusters $\rho_3 = 0.10$, treatment effect heterogeneity at level 2 $\omega_2 = 0.50$, and treatment effect heterogeneity at level 3 $\omega_3 = 0.50$, with level 1 sample size $n = 20$, level 2 sample size $J = 3$, level 3 sample size $K = 30$, with $\alpha = 0.05$ and a two-tailed test. PowerUpR package is used to obtain statistical power values.
Table 5.2 demonstrates that under similar assumptions an RA study can detect smaller MDES values. Contrary to statistical power, the difference is less pronounced where the treatment split is not balanced ($P = 0.33$, and $P = 0.10$). As it is the case with the Table 5.1, to make up for the loss resulting from the correlation between the treatment and assignment variables in RD studies exploratory power of the covariate(s) should be improved significantly. Note that for a balanced split for the treatment ($P = 0.50$) MDES values are similar when $R_1^2$ value change from zero to 0.70. The weaker the correlation between the treatment and assignment variables, the more the discrepancy diminishes. The change in the parameters that directly affects this correlation, if favoring a weak correlation, partially diminishes the discrepancy. As it is the case with the Table 5.1, an assignment variable that is distributed uniformly and a treatment variable derived based on the $P = 0.10$ split indicates an evidence in this direction.


Table 5.2 Minimum Detectable Effect Size for BIRD3rr2rr1 Design

<table>
<thead>
<tr>
<th>Distribution of Z</th>
<th>$P = 0.50$</th>
<th></th>
<th></th>
<th>$P = 0.33$</th>
<th></th>
<th></th>
<th>$P = 0.10$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_1$</td>
<td>RD</td>
<td>RA</td>
<td>RD</td>
<td>RA</td>
<td>RD</td>
<td>RA</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0</td>
<td>0.244</td>
<td>0.191</td>
<td>0.244</td>
<td>0.195</td>
<td>0.280</td>
<td>0.244</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.236</td>
<td>0.187</td>
<td>0.237</td>
<td>0.191</td>
<td>0.270</td>
<td>0.237</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.203</td>
<td>0.173</td>
<td>0.204</td>
<td>0.175</td>
<td>0.226</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.185</td>
<td>0.165</td>
<td>0.185</td>
<td>0.167</td>
<td>0.200</td>
<td>0.185</td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>0</td>
<td>0.275</td>
<td>0.191</td>
<td>0.259</td>
<td>0.195</td>
<td>0.270</td>
<td>0.244</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.265</td>
<td>0.187</td>
<td>0.251</td>
<td>0.191</td>
<td>0.261</td>
<td>0.237</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.222</td>
<td>0.173</td>
<td>0.213</td>
<td>0.175</td>
<td>0.220</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.198</td>
<td>0.165</td>
<td>0.191</td>
<td>0.167</td>
<td>0.196</td>
<td>0.185</td>
<td></td>
</tr>
</tbody>
</table>

Note. Z: Assignment variable. $P$: Proportion of level 1 units in treatment condition. $R^2_1$: Proportion of level 1 variance explained by covariates. RD: Regression Discontinuity. RA: Random Assignment. Values are based on a statistical power $= 0.80$, proportion of variance in the outcome that is explained by level 2 clusters $\rho_2 = 0.20$, proportion of variance in the outcome that is explained by level 3 clusters $\rho_3 = 0.10$, treatment effect heterogeneity for level 2 $\omega_2 = 0.50$, and treatment effect heterogeneity for level 3 $\omega_3 = 0.50$, with level 1 sample size $n = 20$, level 2 sample size $J = 3$, level 3 sample size $K = 30$, with $\alpha = 0.05$ and a two-tailed test. Assignment variable is assumed to follow a normal distribution. PowerUpR package is used to obtain MDES values.
Table 5.3 demonstrates that under similar assumptions an RA study requires much less top-level sample sizes, the major determinant of power. The difference is more pronounced where the treatment split is balanced ($P = 0.50$), however, when the split is not balanced ($P = 0.33$, and $P = 0.10$) and the assignment variable is uniformly distributed, the discrepancy between the two designs the discrepancy partially diminishes. To make up for the loss resulting from the correlation between the treatment and assignment variables in RD studies, exploratory power of the covariate(s) should be improved significantly. Note that for a balanced split for the treatment ($P = 0.50$) required top-level sample sizes are similar when $R^2_1$ value change from zero to 0.70.

Moreover, efficiency values confirms findings in tables above (efficiency is defined as the ratio of the sample size required for RD design and RA design). With a normally distributed assignment variable, RD designs are much less efficient where the treatment split is not balanced ($P = 0.33$, and $P = 0.10$) and where exploratory power of the level 1 covariate is week. However, the efficiency in RD studies tends to slightly increase when the assignment variable is uniformly distributed and the treatment split is not balanced ($P = 0.33$, and $P = 0.10$).
Table 5.3 *Number of Top-level Sample Size for BIRD3rr2rr1 Design*

<table>
<thead>
<tr>
<th>Distribution of Z</th>
<th>$P = 0.50$</th>
<th></th>
<th></th>
<th>$P = 0.33$</th>
<th></th>
<th></th>
<th>$P = 0.10$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_1$</td>
<td>RD</td>
<td>RA</td>
<td>Efficiency</td>
<td>RD</td>
<td>RA</td>
<td>Efficiency</td>
<td>RD</td>
<td>RA</td>
</tr>
<tr>
<td>Normal</td>
<td>0</td>
<td>44</td>
<td>27</td>
<td>1.63</td>
<td>44</td>
<td>29</td>
<td>1.52</td>
<td>57</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>41</td>
<td>27</td>
<td>1.52</td>
<td>41</td>
<td>28</td>
<td>1.46</td>
<td>53</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>31</td>
<td>23</td>
<td>1.35</td>
<td>31</td>
<td>24</td>
<td>1.29</td>
<td>38</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>26</td>
<td>21</td>
<td>1.24</td>
<td>26</td>
<td>21</td>
<td>1.24</td>
<td>30</td>
<td>26</td>
</tr>
<tr>
<td>Uniform</td>
<td>0</td>
<td>55</td>
<td>27</td>
<td>2.04</td>
<td>49</td>
<td>29</td>
<td>1.69</td>
<td>53</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>51</td>
<td>27</td>
<td>1.89</td>
<td>46</td>
<td>28</td>
<td>1.64</td>
<td>50</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>37</td>
<td>23</td>
<td>1.61</td>
<td>34</td>
<td>24</td>
<td>1.42</td>
<td>36</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>29</td>
<td>21</td>
<td>1.38</td>
<td>28</td>
<td>21</td>
<td>1.33</td>
<td>29</td>
<td>26</td>
</tr>
</tbody>
</table>

*Note.* $Z$: Assignment variable. $P$: Proportion of level 1 units in treatment condition. $R^2_1$: Proportion of level 1 variance explained by covariates. RD: Regression Discontinuity. RA: Random Assignment. Values are based on statistical power = 0.80, MDES = 0.20, proportion of variance in the outcome that is explained by level 2 clusters $\rho_2 = 0.20$, proportion of variance in the outcome that is explained by level 3 clusters $\rho_3 = 0.10$, treatment effect heterogeneity at level 2 $\omega_2 = 0.50$, and treatment effect heterogeneity at level 3 $\omega_3 = 0.50$, with level 1 sample size $n = 20$, level 2 sample size $J = 3$, with $\alpha = 0.05$ and a two-tailed test. PowerUpR package is used to obtain number of top-level sample sizes.
Top-level sample size values in Table 5.4 pertains to the design where intercept varies randomly across level 2 and level 3 units but treatment slope is constant. The table compares top-level sample sizes required for regression discontinuity (RD) and random assignment (RA) studies. Table 5.4 demonstrates that under similar assumptions an RA study requires much less top-level sample sizes. The difference is more pronounced where the treatment split is balanced \((P = 0.50)\). To make up for the loss resulting from correlation between the treatment and assignment variables in RD studies exploratory power of the covariate(s) should be improved significantly by including additional variables. Note that for a balanced split for the treatment \((P = 0.50)\) required top-level sample sizes are similar when \(R^2\) value change from zero to 0.70.

Contrary to BIRD3rr2rr1 design, compared to RA studies, RD studies are much less efficient. Efficiency loss compared to RA studies is more pronounced where the treatment split is balanced \((P = 0.50)\), the assignment variable is uniformly distributed, and exploratory power of the level 1 covariate is week. For simple RD designs and for RD designs where the discontinuity is at top-level the efficiency is around 2.75 (balanced treatment split, assuming normally distributed assignment variable). In this regard, BRD3rr2rr1 design is more efficient, and efficiency of BIRD3rc2rc1 design can be similar to the simple RD or where discontinuity reside at top-level when the assignment variable is normally distributed. However, it can be worse when the treatment split is balanced and the assignment variable is uniformly distributed as correspondingly correlation between the treatment and assignment variable become weaker.
### Table 5.4 Number of Top-level Sample Size for BIRD3rc2rc1 Design

<table>
<thead>
<tr>
<th>Distribution of $Z$</th>
<th>$R^2_1$</th>
<th>RD</th>
<th>RA</th>
<th>Efficiency</th>
<th>$P = 0.50$</th>
<th>$P = 0.33$</th>
<th>$P = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>25</td>
<td>9</td>
<td>2.78</td>
<td>25</td>
<td>10</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>23</td>
<td>8</td>
<td>2.88</td>
<td>23</td>
<td>9</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>13</td>
<td>5</td>
<td>2.60</td>
<td>13</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>8</td>
<td>3</td>
<td>2.67</td>
<td>8</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Uniform</td>
<td>0</td>
<td>37</td>
<td>9</td>
<td>4.11</td>
<td>31</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>33</td>
<td>8</td>
<td>4.13</td>
<td>28</td>
<td>9</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>18</td>
<td>5</td>
<td>3.60</td>
<td>15</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>11</td>
<td>3</td>
<td>3.67</td>
<td>9</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

*Note.* $Z$: Assignment variable. $P$: Proportion of level 1 units in treatment condition. $R^2_1$: Proportion of level 1 variance explained by covariates. RD: Regression Discontinuity. RA: Random Assignment. Values are based on statistical power $= 0.80$, MDES $= 0.20$, proportion of variance in the outcome that is explained by level 2 clusters $\rho_2 = 0.20$, proportion of variance in the outcome that is explained by level 3 clusters $\rho_3 = 0.10$, with level 1 sample size $n = 20$, level 2 sample size $J = 3$, with $\alpha = 0.05$ and a two-tailed test. PowerUpR package is used to obtain number of top-level sample sizes.
Since there does not exist a formula to design a three-level RD study where treatment and assignment variable resides at level 1, researcher may be tempted to use a two-level RD study and use parameters from misspecified models where either intermediate- or top-level is ignored. From an analysis perspective, when intermediate-level is ignored, variance of the ignored level shifts to the new bottom and to the new top level. This result in slightly underestimation of power and can be neglected if the variance at the intermediate-level is small to begin with. However, for example, classroom level variance can exceed school level variance in practice (Goldstein, 2011; Muthen, 1991). If the reason to ignore an intermediate-level is based upon small sample sizes, statistical techniques such as including the intermediate-level as fixed effects (Van Den Noorthgate, Opdenakker, & Onghena, 2005) or bootstrapping (Goldstein, 2011) can help mitigate the problem.

Ignoring the top-level is more problematic despite the variance component at the third level being small. When top-level is ignored, variance of the ignored level shifts to the new top level. This results in overestimation power. Compared to ignoring-intermediate-level the distortion in ignoring the top-level is more pronounced as the top-level sample size change dramatically. On its own, this would not constitute a problem, however, Type I error rates are also inflated. This means research more likely than usual to detect a treatment effect when there is no treatment effect.

Finally, calculation of sample sizes is not straight forward as units or clusters are often associated with certain cost. This study proposed a general framework for constrained optimal sample allocation that is applicable to both multilevel RD and multilevel RA studies. The framework allows unequal marginal cost specifications for
treatment and control conditions, therefore, allows optimization of $P$ in RA design in addition to sample sizes. Although unequal marginal cost specification is also applicable to RD designs, the framework does not allow optimization of $P$ as one of the key assumption in RD studies is $P$ is derived based on a predetermined cutoff value. The framework allow researcher to fix sample sizes for one or more levels or place bound constraints to find optimal number of subjects or clusters in addition to budget constraints.

As number of parameters to be optimized increases, finding a local solution is more likely. Therefore, solution from multiple algorithms should be compared. Appendix B provides tables that compares efficiency of four algorithms to find COSA solutions. Efficiency of an algorithm is measured to the extent the power is superior to other algorithms under budget constraint and the total cost does not deviate from the budget significantly. Although results are limited to three-level cluster RA and RD studies, and other design parameters, it provide a glance on the behavior of the algorithms. Algorithms are compared by varying starting values and marginal cost per units while placing constraints on one level at a time. When only one parameter is optimized, all algorithms produce same solutions, when two parameters are optimized LBFGS algorithm produces superior results. The second best algorithm is SQSLS which is the default in PowerUpR. Although LBFGS algorithm produces best solutions in the contexts that were defined, its failure to converge from time to time under power and MDES constraints that makes it less ideal to assign it as a default. On the other hand, starting values play an important role on the solution especially when two or more parameters are optimized. It is recommended that users use various starting values and compare different
algorithms for an optimal solution. Placing bound constraints is an efficient way of doing this because starting values are overridden by the averages from bounds.

Table 5.5 compares COSA solutions in Optimal Design Plus (OD+) and PowerUpR for a two-level cluster randomized trial. Two-level cluster randomized trial was chosen merely because OD+ provides a COSA solution only for this design. This is a specific case of RD design where the relationship between treatment and assignment variables is zero. This can be achieved in PowerUpR by specifying RTZ = 0. Results indicate some important distinction between the two packages. PowerUpR package does not assume sample size for lower levels are integers as these values are averages or harmonic means, therefore, under reasonable starting values or bound constraints it provides more precise solution than OD+. However, if a researcher prefers integers to avoid confusion, sample size values for lower levels can be rounded and specified as fixed parameters in a second run.
Table 5.5 Comparison of COSA Solutions in PowerUpR and Optimal Design Plus (OD+) for a Two-level Cluster Randomized Controlled Trial

<table>
<thead>
<tr>
<th>PowerUpR</th>
<th>OD+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Optimal Solution</strong></td>
<td><strong>Starting Values</strong></td>
</tr>
<tr>
<td>n</td>
<td>J</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Budget = $10,000</strong></td>
<td></td>
</tr>
<tr>
<td>$6.32$</td>
<td>$31$</td>
</tr>
<tr>
<td>$6.32$</td>
<td>$31$</td>
</tr>
<tr>
<td>$6.32$</td>
<td>$31$</td>
</tr>
<tr>
<td><strong>Budget = $50,000$</strong></td>
<td></td>
</tr>
<tr>
<td>$7.69$</td>
<td>$141$</td>
</tr>
<tr>
<td>$6.33$</td>
<td>$153$</td>
</tr>
<tr>
<td>$15$</td>
<td>$100$</td>
</tr>
<tr>
<td>$6.33$</td>
<td>$153$</td>
</tr>
<tr>
<td><strong>Budget = $100,000$</strong></td>
<td></td>
</tr>
<tr>
<td>$22.83$</td>
<td>$152$</td>
</tr>
<tr>
<td>$25.43$</td>
<td>$141$</td>
</tr>
<tr>
<td>$15$</td>
<td>$100$</td>
</tr>
<tr>
<td>$6.32$</td>
<td>$306$</td>
</tr>
</tbody>
</table>

*Note. Other parameters used in PowerUpR: $cn = 20$, $cJ = 200$, constrain = "cost", alpha = 0.05, two.tail = TRUE, mdes = 0.25, optimizer = "auglag_slsqp", rho2 = 0.20, P = 0.50, g2 = 0, R12 = 0, R22 = 0.*
Although top-level sample size is a major determinant of power, the ramification of small sample sizes at level 1 and level 2 may be different for formula than it is in reality. In reality estimated variance components associated with the lower level small sample sizes are not stable and affects statistical power and Type I error rates. Researcher should be aware that the formula assumes a balanced design and correctly specified models. Significant deviation from a balanced design and parameters used from misspecified models may produce unwanted results. Researcher should also be aware of the relationship between treatment split (P in PowerUpR) and correlation between the treatment and assignment variable (RTZ in PowerUpR). P and RTZ are not independent from each other, therefore, researcher should be cautious to specify RTZ. The default is NULL and the RTZ is calculated from the specified distribution based on P in PowerUpR, but it is included to provide the research the flexibility to consider various other distributions for the assignment variable.

Placing bound constraint to find COSA solutions can be both a powerful aspect of the COSA framework but it can also be a limitation depending on familiarity of the researcher with the statistical power analysis. For example, as seen in Table 5.5, underestimating upper-bounds limits the sample size search and automatically picks upper bounds. However, the total cost significantly deviates from the budget, therefore, researcher can alter upper-bounds accordingly.

Moreover, another limitation comes from optimization of treatment split in RA studies, as often, complementary proportions (sum of which equals to unity) produces same results. A failure to converge on a global solution will result in a treatment split that might be efficient economically but not effort wise. For example, a treatment split at the
$P = 70$ requires more effort whereas $P = 30$ could be both efficient and require less effort during implementation. Researcher can alter default values accordingly if this is suspected.

With optimization of multiple parameters, there are cases where global solution is not guaranteed (mostly four algorithms will not agree on a common solution), in this case, using integer approximations along with proper bound constraints can provide researcher guidance. In a second step, starting values can be placed with the best integer solution. This will improve the integer solution to the decimal points.

Another drawback is navigating PowerUpR package in R can be a cumbersome procedure for some researcher. In the meantime, a web page is tentatively provided for research who wish to use COSA functions by menu and click fashion. The interface and where the web-page is hosted may change in the future, therefore it is recommended that researcher frequently check updates from www.causalevalution.org website.

Finally, numerical optimization allows flexibility with finding COSA solution. A possible direction for COSA researcher is to adjust COSA solution for multilevel attrition and to add more algorithms that have better global convergence properties such as simulated annealing.
References


Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects?


Ypma, J. (2014). nloptr: R interface to NLopt. R package version 1.0.4

## Table A.1.1

### Power Analysis for BIRD3rc2rc1 Design

| Scenario | \( \hat{\xi}_{100} \) | SE(\( \hat{\xi}_{100} \)) | ES(\( \hat{\xi}_{100} \)) | \( \rho_2 \) | \( \rho_3 \) | \( R^2_1 \) | \( P \) | \( \rho_{TZ} \) | \( K \) | SE_emp(\( \hat{\xi}_{100} \)) | SE_eq(\( \hat{\xi}_{100} \)) | AD in SEs | RD in SEs (%) | 95% CI Coverage | Empirical Power | Calculated Power | AD in Powers | RD in Powers (%) |
|----------|-----------------|-----------------|-----------------|-----------|-----------|----------|------|----------|------|-----------------|-----------------|-------------|--------------|----------------|-----------------|-----------------|-----------------|--------------|-----------------|
| P1       | 0.30            | 0.19            | 0.35            | 0.11     | 0.47     | 0.50    | 0.80 | 0.08     | 8    | 0.19            | 0.19            | 0.00        | 1.36         | 0.95          | 0.35           | 0.35           | 0.00         | -1.05          |
| P2       | 0.30            | 0.14            | 0.35            | 0.11     | 0.47     | 0.50    | 0.80 | 0.08     | 15   | 0.14            | 0.14            | 0.00        | 1.27         | 0.95          | 0.58           | 0.58           | 0.00         | -0.30          |
| P3       | 0.30            | 0.09            | 0.35            | 0.11     | 0.47     | 0.50    | 0.80 | 0.08     | 35   | 0.09            | 0.09            | 0.00        | -0.85        | 0.95          | 0.91           | 0.90           | 0.00         | -0.28          |
| P4       | 0.30            | 0.20            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 8    | 0.20            | 0.20            | 0.00        | -1.23        | 0.95          | 0.32           | 0.32           | 0.01         | 1.00           |
| P5       | 0.30            | 0.15            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 15   | 0.15            | 0.15            | 0.01        | -0.99        | 0.96          | 0.86           | 0.86           | 0.01         | 1.61           |
| P6       | 0.30            | 0.10            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 35   | 0.10            | 0.10            | 0.00        | -3.75        | 0.96          | 0.63           | 0.63           | 0.00         | -10.49         |
| P7       | 0.30            | 0.10            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 8    | 0.10            | 0.10            | 0.00        | -1.61        | 0.96          | 0.34           | 0.34           | 0.01         | -8.46          |
| P8       | 0.30            | 0.10            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 15   | 0.10            | 0.10            | 0.00        | -1.33        | 0.96          | 0.34           | 0.34           | 0.01         | -2.40          |
| P9       | 0.30            | 0.10            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 35   | 0.10            | 0.10            | 0.00        | -0.33        | 0.96          | 0.34           | 0.34           | 0.01         | -1.06          |
| P10      | 0.30            | 0.10            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 8    | 0.10            | 0.10            | 0.00        | -0.26        | 0.96          | 0.34           | 0.34           | 0.00         | 0.87           |
| P11      | 0.30            | 0.10            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 15   | 0.10            | 0.10            | 0.00        | -0.26        | 0.96          | 0.34           | 0.34           | 0.00         | 0.87           |
| P12      | 0.30            | 0.10            | 0.36            | 0.11     | 0.46     | 0.46    | 0.70 | 0.70     | 35   | 0.10            | 0.10            | 0.00        | -0.26        | 0.96          | 0.34           | 0.34           | 0.00         | 0.87           |

**Note.** Results are based on 5000 replications. \( \hat{\xi}_{100} \): Treatment effect. SE: Standard Error. ES: Effect size. \( \rho_2 \): Proportion of variance in the outcome between level 2 units. \( \rho_3 \): Proportion of variance in the outcome between level 3 units. \( R^2_1 \): Proportion of variance in the outcome explained level 1 covariates. \( P \): Proportion of subjects fall below (or above) cutoff score on the assignment variable. \( \rho_{TZ} \): Correlation between the assignment variable and the treatment status. \( n \): Average number of level 1 units per level 2 units, which is set to 30. \( J \): Average number of level 2 units per level 3 units, which is set to 5. \( K \): Number of level 3 units. AD: Absolute difference. RD: Relative difference.
### Table A.1.2

**Type I Error Analysis for BIRD3rc2rc1 Design**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
<th>T9</th>
<th>T10</th>
<th>T11</th>
<th>T12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\xi}_{100}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$SE(\hat{\xi}_{100})$</td>
<td>0.19</td>
<td>0.14</td>
<td>0.09</td>
<td>0.20</td>
<td>0.15</td>
<td>0.10</td>
<td>0.18</td>
<td>0.13</td>
<td>0.09</td>
<td>0.20</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>$ES(\hat{\xi}_{100})$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.16</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>$R_i^2$</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>$\rho_{TZ}$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$K$</td>
<td>8</td>
<td>15</td>
<td>35</td>
<td>8</td>
<td>15</td>
<td>35</td>
<td>8</td>
<td>15</td>
<td>35</td>
<td>8</td>
<td>15</td>
<td>35</td>
</tr>
</tbody>
</table>

| $SE_{emp}(\hat{\xi}_{100})$ | 0.19 | 0.14 | 0.09 | 0.21 | 0.15 | 0.10 | 0.18 | 0.13 | 0.09 | 0.20 | 0.15 | 0.10 |
| $SE_{eq}(\hat{\xi}_{100})$ | 0.19 | 0.14 | 0.09 | 0.20 | 0.15 | 0.10 | 0.19 | 0.14 | 0.09 | 0.20 | 0.15 | 0.10 |
| AD in SEs | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| RD in SEs (%) | 0.94 | 1.04 | -0.92 | -1.47 | -1.72 | -2.33 | 3.74 | 4.13 | 4.66 | 0.44 | -0.95 | 0.08 |

| 95% CI Coverage | 0.95 | 0.95 | 0.94 | 0.94 | 0.95 | 0.96 | 0.96 | 0.96 | 0.95 | 0.95 | 0.95 | 0.95 |
| Empirical Type I Error | 0.05 | 0.05 | 0.06 | 0.05 | 0.05 | 0.05 | 0.06 | 0.06 | 0.05 | 0.05 | 0.05 | 0.05 |
| Calculated Type I Error | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| AD in Type I Errors | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 | -0.01 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| RD in Type I Errors (%) | -2.33 | -5.62 | -13.39 | -6.34 | 0.11 | -3.09 | -9.42 | -9.59 | -12.55 | 0.40 | -4.58 | -1.15 |

*Note.* Results are based on 5000 replications. $\hat{\xi}_{100}$: Treatment effect. $SE$: Standard Error. $ES$: Effect size. $\rho_2$: Proportion of variance in the outcome between level 2 units. $\rho_3$: Proportion of variance in the outcome between level 3 units. $R_i^2$: Proportion of variance in the outcome explained level 1 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $\rho_{TZ}$: Correlation between the assignment variable and the treatment status. $n$: Average number of level 1 units per level 2 units, which is set to 30. $J$: Average number of level 2 units per level 3 units, which is set to 5. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
### Table A.1.3

**Power Analysis for BIRD3fc2rr1 Design**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
<th>P11</th>
<th>P12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_{100}$</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>$SE(\xi_{100})$</td>
<td>0.12</td>
<td>0.17</td>
<td>0.24</td>
<td>0.12</td>
<td>0.18</td>
<td>0.25</td>
<td>0.11</td>
<td>0.16</td>
<td>0.23</td>
<td>0.12</td>
<td>0.17</td>
<td>0.24</td>
</tr>
<tr>
<td>$ES(\xi_{100})$</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.24</td>
<td>0.23</td>
<td>0.22</td>
<td>0.22</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>$\omega^2$</td>
<td>0.36</td>
<td>0.37</td>
<td>0.39</td>
<td>0.37</td>
<td>0.39</td>
<td>0.41</td>
<td>0.55</td>
<td>0.55</td>
<td>0.56</td>
<td>0.15</td>
<td>0.20</td>
<td>0.27</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>$R^2_T$</td>
<td>0.34</td>
<td>0.34</td>
<td>0.35</td>
<td>0.26</td>
<td>0.28</td>
<td>0.33</td>
<td>0.62</td>
<td>0.60</td>
<td>0.61</td>
<td>0.17</td>
<td>0.26</td>
<td>0.38</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$\rho_TZ$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$K$</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

| $SE_{emp}(\xi_{100})$ | 0.12 | 0.17 | 0.24 | 0.12 | 0.18 | 0.25 | 0.11 | 0.16 | 0.22 | 0.12 | 0.17 | 0.24 |
| $SE_{eq}(\xi_{100})$ | 0.12 | 0.16 | 0.23 | 0.12 | 0.17 | 0.24 | 0.11 | 0.16 | 0.22 | 0.12 | 0.16 | 0.23 |
| AD in SEs | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 | 0.00 | -0.01 | -0.01 | -0.01 |
| 95% CI Coverage | 0.94 | 0.95 | 0.94 | 0.95 | 0.94 | 0.95 | 0.94 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 |
| Empirical Power | 0.84 | 0.56 | 0.31 | 0.80 | 0.50 | 0.29 | 0.87 | 0.59 | 0.32 | 0.83 | 0.54 | 0.30 |
| Calculated Power | 0.85 | 0.57 | 0.33 | 0.82 | 0.52 | 0.31 | 0.88 | 0.60 | 0.33 | 0.86 | 0.57 | 0.33 |
| AD in Powers | 0.02 | 0.01 | 0.01 | 0.02 | 0.02 | 0.01 | 0.00 | 0.01 | 0.03 | 0.03 | 0.03 | 0.03 |
| RD in Powers (%) | 1.88 | 1.34 | 4.21 | 2.44 | 3.84 | 5.25 | 1.05 | 0.53 | 4.45 | 3.80 | 5.92 | 8.32 |

*Note.* Results are based on 5000 replications. $\xi_{100}$: Treatment effect. SE: Standard Error, ES: Effect size. $\rho_2$: Proportion of variance in the outcome between level 2 units. $\omega^2$: Treatment effect heterogeneity across level 2 units. $R^2_T$: Proportion of variance in the outcome explained level 1 covariates. $R^2_{2T}$: Proportion of variance in the treatment effect explained level 2 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $\rho_TZ$: Correlation between the assignment variable and the treatment status. $n$: Average number of level 1 units per level 2 units, which is set to 20. $J$: Average number of level 2 units per level 3 units, which is set to 5. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
### Table A.1.4

**Type I Error Analysis for BIRD3fc2rr1 Design**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
<th>T9</th>
<th>T10</th>
<th>T11</th>
<th>T12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_{100}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$SE(\xi_{100})$</td>
<td>0.12</td>
<td>0.17</td>
<td>0.24</td>
<td>0.12</td>
<td>0.18</td>
<td>0.25</td>
<td>0.11</td>
<td>0.16</td>
<td>0.23</td>
<td>0.12</td>
<td>0.17</td>
<td>0.24</td>
</tr>
<tr>
<td>$ES(\xi_{100})$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.26</td>
<td>0.26</td>
<td>0.25</td>
<td>0.23</td>
<td>0.23</td>
<td>0.22</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.36</td>
<td>0.37</td>
<td>0.40</td>
<td>0.38</td>
<td>0.39</td>
<td>0.41</td>
<td>0.55</td>
<td>0.56</td>
<td>0.57</td>
<td>0.16</td>
<td>0.20</td>
<td>0.27</td>
</tr>
<tr>
<td>$R^2_1$</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>$R^2_{2T}$</td>
<td>0.34</td>
<td>0.34</td>
<td>0.36</td>
<td>0.26</td>
<td>0.28</td>
<td>0.32</td>
<td>0.62</td>
<td>0.61</td>
<td>0.61</td>
<td>0.17</td>
<td>0.27</td>
<td>0.38</td>
</tr>
<tr>
<td>$\rho_{TZ}$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$K$</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

| $SE_{emp}(\xi_{100})$ | 0.12 | 0.17 | 0.24 | 0.13 | 0.18 | 0.25 | 0.11 | 0.16 | 0.23 | 0.12 | 0.17 | 0.24 |
| $SE_{eq}(\xi_{100})$ | 0.12 | 0.16 | 0.23 | 0.12 | 0.17 | 0.24 | 0.11 | 0.16 | 0.22 | 0.12 | 0.16 | 0.23 |
| AD in SEs | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 |
| RD in SEs (%) | -2.84 | -0.48 | -1.63 | -3.14 | -2.24 | -2.67 | -2.28 | -2.96 | -1.76 | -1.63 | -2.26 | -2.29 |
| Empirical Type I Error | 0.94 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 | 0.95 | 0.94 | 0.95 | 0.95 | 0.94 |
| Calculated Type I Error | 0.94 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 | 0.95 | 0.94 | 0.95 | 0.95 | 0.94 |
| AD in Type I Errors | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| RD in Type I Errors (%) | -6.72 | 14.88 | 1.68 | -3.47 | -1.36 | -0.32 | 0.00 | -6.01 | -3.85 | 5.59 | 7.81 | 1.65 |

*Note.* Results are based on 5000 replications. $\xi_{100}$: Treatment effect. $SE$: Standard Error. $ES$: Effect size. $\rho_2$: Proportion of variance in the outcome between level 2 units. $\omega_2$: Treatment effect heterogeneity across level 2 units. $R^2_1$: Proportion of variance in the outcome explained level 1 covariates. $R^2_{2T}$: Proportion of variance in the treatment effect explained level 2 covariates. $\rho_{TZ}$: Correlation between the assignment variable and the treatment status. $n$: Average number of level 1 units per level 2 units, which is set to 20. $J$: Average number of level 2 units per level 3 units, which is set to 5. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
<th>P11</th>
<th>P12</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\xi_{100})</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>(SE(\xi_{100}))</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>(ES(\xi_{100}))</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>(\rho_2)</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>(R_1^2)</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>(P)</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>(\rho_{TZ})</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>(K)</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

| \(SE_{emp}(\xi_{100})\) | 0.08 | 0.11 | 0.16 | 0.08 | 0.12 | 0.17 | 0.08 | 0.11 | 0.16 | 0.08 | 0.12 | 0.17 |
| \(SE_{eq}(\xi_{100})\) | 0.08 | 0.11 | 0.16 | 0.08 | 0.12 | 0.17 | 0.08 | 0.11 | 0.16 | 0.08 | 0.12 | 0.17 |
| AD in SEs | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 |
| RD in SEs (%) | -0.90 | -2.81 | -1.77 | -2.05 | -1.91 | -1.32 | -1.56 | -2.12 | -2.76 | -1.74 | -2.63 | -3.46 |

| 95% CI Coverage | 0.95 | 0.94 | 0.95 | 0.94 | 0.95 | 0.95 | 0.95 | 0.95 | 0.94 | 0.95 | 0.94 | 0.94 |
| Empirical Power | 0.88 | 0.60 | 0.34 | 0.84 | 0.55 | 0.32 | 0.88 | 0.59 | 0.33 | 0.84 | 0.56 | 0.31 |
| Calculated Power | 0.89 | 0.62 | 0.35 | 0.85 | 0.56 | 0.33 | 0.89 | 0.62 | 0.35 | 0.85 | 0.58 | 0.32 |
| AD in Powers | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.01 | 0.02 | 0.01 |
| RD in Powers (%) | 0.86 | 2.45 | 2.94 | 0.85 | 2.61 | 2.92 | 1.20 | 3.76 | 4.71 | 1.06 | 3.49 | 1.84 |

*Note.* Results are based on 5000 replications. \(\xi_{100}\): Treatment effect. SE: Standard Error. ES: Effect size. \(\rho_2\): Proportion of variance in the outcome between level 2 units. \(R_1^2\): Proportion of variance in the outcome explained level 1 covariates. \(P\): Proportion of subjects fall below (or above) cutoff score on the assignment variable. \(\rho_{TZ}\): Correlation between the assignment variable and the treatment status. \(n\): Average number of level 1 units per level 2 units, which is set to 20. \(J\): Average number of level 2 units per level 3 units, which is set to 5. \(K\): Number of level 3 units. AD: Absolute difference. RD: Relative difference.
Table A.1.6

Type I Error Analysis for BIRD3fc2rc1 Design

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
<th>T9</th>
<th>T10</th>
<th>T11</th>
<th>T12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{\xi}_{100}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$SE(\tilde{\xi}_{100})$</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>$ES(\tilde{\xi}_{100})$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>$R^2_1$</td>
<td>0.47</td>
<td>0.47</td>
<td>0.47</td>
<td>0.47</td>
<td>0.47</td>
<td>0.47</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$P_{TZ}$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$K$</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>$SE_{emp}(\tilde{\xi}_{100})$</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>$SE_{eq}(\tilde{\xi}_{100})$</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
<td>0.08</td>
<td>0.11</td>
<td>0.16</td>
<td>0.08</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>AD in SEs</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RD in SEs (%)</td>
<td>-2.19</td>
<td>-1.68</td>
<td>-2.16</td>
<td>-2.28</td>
<td>-3.43</td>
<td>-2.33</td>
<td>-1.27</td>
<td>-2.10</td>
<td>-1.39</td>
<td>-2.23</td>
<td>-3.75</td>
<td>-4.00</td>
</tr>
<tr>
<td>95% CI Coverage</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Empirical Type I Error</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Calculated Type I Error</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>AD in Type I Errors</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RD in Type I Errors (%)</td>
<td>0.81</td>
<td>6.88</td>
<td>0.45</td>
<td>-1.88</td>
<td>5.20</td>
<td>-5.30</td>
<td>2.88</td>
<td>2.46</td>
<td>8.83</td>
<td>2.88</td>
<td>-4.57</td>
<td>4.63</td>
</tr>
</tbody>
</table>

Note. Results are based on 5000 replications. $\tilde{\xi}_{100}$: Treatment effect. SE: Standard Error. ES: Effect size. $\rho_2$: Proportion of variance in the outcome between level 2 units. $R^2_1$: Proportion of variance in the outcome explained level 1 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $P_{TZ}$: Correlation between the assignment variable and the treatment status. $n$: Average number of level 1 units per level 2 units, which is set to 20. $J$: Average number of level 2 units per level 3 units, which is set to 5. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
### Table A.1.7

*Power Analysis for BIRD3fc2fc1 Design*

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\xi}_{100}$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>SE($\hat{\xi}_{100}$)</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
<td>0.11</td>
<td>0.10</td>
<td>0.16</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>ES($\hat{\xi}_{100}$)</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>$R^2_1$</td>
<td>0.81</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.48</td>
<td>0.47</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$\rho_{TZ}$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$K$</td>
<td>25</td>
<td>10</td>
<td>25</td>
<td>10</td>
<td>25</td>
<td>10</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>SE$<em>{emp}$($\hat{\xi}</em>{100}$)</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
<td>0.11</td>
<td>0.10</td>
<td>0.16</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>SE$<em>{eq}$($\hat{\xi}</em>{100}$)</td>
<td>0.07</td>
<td>0.10</td>
<td>0.07</td>
<td>0.11</td>
<td>0.10</td>
<td>0.16</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>AD in SEs</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RD in SEs (%)</td>
<td>-2.61</td>
<td>-1.82</td>
<td>-3.52</td>
<td>-2.52</td>
<td>-3.01</td>
<td>-3.20</td>
<td>-3.78</td>
<td>-2.62</td>
</tr>
<tr>
<td>95% CI Coverage</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Empirical Power</td>
<td>0.96</td>
<td>0.64</td>
<td>0.94</td>
<td>0.61</td>
<td>0.68</td>
<td>0.35</td>
<td>0.65</td>
<td>0.30</td>
</tr>
<tr>
<td>Calculated Power</td>
<td>0.96</td>
<td>0.67</td>
<td>0.95</td>
<td>0.62</td>
<td>0.71</td>
<td>0.35</td>
<td>0.67</td>
<td>0.32</td>
</tr>
<tr>
<td>AD in Powers</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>RD in Powers (%)</td>
<td>0.68</td>
<td>3.50</td>
<td>1.17</td>
<td>2.07</td>
<td>3.81</td>
<td>1.68</td>
<td>2.10</td>
<td>5.57</td>
</tr>
</tbody>
</table>

*Note.* Results are based on 5000 replications. $\hat{\xi}_{100}$: Treatment effect. SE: Standard Error. ES: Effect size. $R^2_1$: Proportion of variance in the outcome explained level 1 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $\rho_{TZ}$: Correlation between the assignment variable and the treatment status. $n$: Average number of level 1 units per level 2 units, which is set to 20. $J$: Average number of level 2 units per level 3 units, which is set to 5. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
Table A.1.8

Type I Error Analysis for BIRD3fc2fc1 Design

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\xi}_{100}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$SE(\hat{\xi}_{100})$</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
<td>0.11</td>
<td>0.10</td>
<td>0.16</td>
<td>0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>$ES(\hat{\xi}_{100})$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$R^2_1$</td>
<td>0.80</td>
<td>0.79</td>
<td>0.80</td>
<td>0.79</td>
<td>0.48</td>
<td>0.46</td>
<td>0.48</td>
<td>0.46</td>
</tr>
<tr>
<td>$P$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
<td>0.50</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$\rho_{TZ}$</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.80</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$K$</td>
<td>25</td>
<td>10</td>
<td>25</td>
<td>10</td>
<td>25</td>
<td>10</td>
<td>25</td>
<td>10</td>
</tr>
</tbody>
</table>

| $SE_{emp}(\hat{\xi}_{100})$ | 0.07 | 0.11 | 0.07 | 0.11 | 0.10 | 0.16 | 0.11 | 0.17 |
| $SE_{eq}(\hat{\xi}_{100})$ | 0.07 | 0.10 | 0.07 | 0.11 | 0.10 | 0.16 | 0.11 | 0.17 |
| AD in SEs | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| RD in SEs (%) | -3.10 | -1.60 | -2.42 | -2.42 | -0.64 | -2.50 | -2.38 | -2.83 |
| 95% CI Coverage | 0.95 | 0.95 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 | 0.94 |
| Empirical Type I Error | 0.05 | 0.04 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| Calculated Type I Error | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| AD in Type I Errors | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| RD in Type I Errors (%) | 0.08 | 15.22 | 5.05 | -6.22 | 8.23 | -2.34 | -7.75 | 2.10 |

Note. Results are based on 5000 replications. $\hat{\xi}_{100}$: Treatment effect. SE: Standard Error. ES: Effect size. $R^2_1$: Proportion of variance in the outcome explained level 1 covariates. $P$: Proportion of subjects fall below (or above) cutoff score on the assignment variable. $\rho_{TZ}$: Correlation between the assignment variable and the treatment status. $n$: Average number of level 1 units per level 2 units, which is set to 20. $J$: Average number of level 2 units per level 3 units, which is set to 5. $K$: Number of level 3 units. AD: Absolute difference. RD: Relative difference.
A.2 Ignoring a Level of Nesting

Table A.2.1

*Comparison of Power from the Correctly Specified and the Level 2 Ignored Model*

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical Power from BIRD3rr2rr1 (0)</td>
<td>0.44</td>
<td>0.30</td>
<td>0.74</td>
<td>0.52</td>
<td>0.45</td>
<td>0.26</td>
<td>0.72</td>
<td>0.45</td>
</tr>
<tr>
<td>Empirical Power from BIRD2rr1 (1)</td>
<td>0.38</td>
<td>0.28</td>
<td>0.65</td>
<td>0.49</td>
<td>0.38</td>
<td>0.24</td>
<td>0.62</td>
<td>0.42</td>
</tr>
<tr>
<td>Calculated Power for BIRD2rr1 (2)</td>
<td>0.35</td>
<td>0.24</td>
<td>0.64</td>
<td>0.44</td>
<td>0.36</td>
<td>0.22</td>
<td>0.61</td>
<td>0.41</td>
</tr>
<tr>
<td>AD in Powers (1-0)</td>
<td>-0.07</td>
<td>-0.03</td>
<td>-0.09</td>
<td>-0.03</td>
<td>-0.07</td>
<td>-0.01</td>
<td>-0.10</td>
<td>-0.03</td>
</tr>
<tr>
<td>AD in Powers (2-0)</td>
<td>-0.09</td>
<td>-0.06</td>
<td>-0.10</td>
<td>-0.08</td>
<td>-0.09</td>
<td>-0.03</td>
<td>-0.11</td>
<td>-0.04</td>
</tr>
<tr>
<td>RD in Powers (1-0)</td>
<td>-15.09</td>
<td>-8.29</td>
<td>-12.05</td>
<td>-5.16</td>
<td>-16.02</td>
<td>-5.53</td>
<td>-13.57</td>
<td>-7.28</td>
</tr>
</tbody>
</table>

*Note.* AD: Absolute difference. RD: Relative difference (%).

Table A.2.2

*Comparison of Type I Error from the Correctly Specified and the Level 2 Ignored Model*

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical Type I Error from BIRD3rr2rr1 (0)</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Empirical Type I Error from BIRD2rr1 (1)</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Calculated Type I Error for BIRD2rr1 (2)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>AD in Type I Errors (1-0)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>AD in Type I Errors (2-0)</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RD in Type I Errors (1-0)</td>
<td>-5.90</td>
<td>2.14</td>
<td>8.95</td>
<td>-4.17</td>
<td>2.19</td>
<td>-2.45</td>
<td>-12.04</td>
<td>-8.65</td>
</tr>
<tr>
<td>RD in Type I Errors (2-0)</td>
<td>-13.18</td>
<td>-10.71</td>
<td>-2.70</td>
<td>-5.28</td>
<td>-8.76</td>
<td>-12.55</td>
<td>-8.76</td>
<td>-5.85</td>
</tr>
</tbody>
</table>

*Note.* AD: Absolute difference. RD: Relative difference (%).
Table A.2.3

Comparison of Power from Correctly Specified and Level 3 Ignored Model

<table>
<thead>
<tr>
<th>Scenario</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical Power from BIRD3rr2rr1 (0)</td>
<td>0.44</td>
<td>0.30</td>
<td>0.74</td>
<td>0.52</td>
<td>0.45</td>
<td>0.26</td>
<td>0.72</td>
<td>0.45</td>
</tr>
<tr>
<td>Empirical Power from BIRD2rr1 (1)</td>
<td>0.62</td>
<td>0.43</td>
<td>0.86</td>
<td>0.66</td>
<td>0.63</td>
<td>0.34</td>
<td>0.84</td>
<td>0.54</td>
</tr>
<tr>
<td>Calculated Power for BIRD2rr1 (2)</td>
<td>0.66</td>
<td>0.33</td>
<td>0.93</td>
<td>0.60</td>
<td>0.68</td>
<td>0.31</td>
<td>0.92</td>
<td>0.56</td>
</tr>
<tr>
<td>AD in Powers (1-0)</td>
<td>0.18</td>
<td>0.13</td>
<td>0.12</td>
<td>0.14</td>
<td>0.19</td>
<td>0.08</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>AD in Powers (2-0)</td>
<td>0.22</td>
<td>0.03</td>
<td>0.19</td>
<td>0.08</td>
<td>0.23</td>
<td>0.05</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>RD in Powers (1-0)</td>
<td>40.68</td>
<td>43.44</td>
<td>16.00</td>
<td>26.01</td>
<td>41.50</td>
<td>30.66</td>
<td>16.52</td>
<td>20.36</td>
</tr>
<tr>
<td>RD in Powers (2-0)</td>
<td>49.34</td>
<td>11.00</td>
<td>26.11</td>
<td>15.58</td>
<td>51.11</td>
<td>19.18</td>
<td>28.32</td>
<td>23.27</td>
</tr>
</tbody>
</table>

Note. AD: Absolute difference. RD: Relative difference (%).

Table A.2.4

Comparison of Type I Error from Correctly Specified and Level 3 Ignored Model

<table>
<thead>
<tr>
<th>Scenario</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical Type I Error from BIRD3rr2rr1 (0)</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Empirical Type I Error from BIRD2rr1 (1)</td>
<td>0.15</td>
<td>0.15</td>
<td>0.16</td>
<td>0.14</td>
<td>0.14</td>
<td>0.10</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>Calculated Type I Error for BIRD2rr1 (2)</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>AD in Type I Errors (1-0)</td>
<td>0.09</td>
<td>0.09</td>
<td>0.11</td>
<td>0.09</td>
<td>0.09</td>
<td>0.05</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>AD in Type I Errors (2-0)</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RD in Type I Errors (1-0)</td>
<td>159.03</td>
<td>158.93</td>
<td>208.56</td>
<td>168.94</td>
<td>161.68</td>
<td>79.72</td>
<td>156.20</td>
<td>90.60</td>
</tr>
<tr>
<td>RD in Type I Errors (2-0)</td>
<td>-13.14</td>
<td>-10.71</td>
<td>-2.72</td>
<td>-5.27</td>
<td>-8.70</td>
<td>-12.52</td>
<td>-8.76</td>
<td>-5.58</td>
</tr>
</tbody>
</table>

Note. AD: Absolute difference. RD: Relative difference (%).
Appendix B

Optimizer Comparison in COSA

B.1 Three level Cluster-level Regression Discontinuity (CRD) with Discontinuity at Level 3

Table B.1.1

*Three-level CRD design: Equal marginal cost, starting values for n, J, K is (30, 3, 100)*

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer (AUGLAG +)</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>36.69</td>
<td>2.22</td>
<td>78</td>
<td>0.50</td>
<td>$100,280.00</td>
<td>0.31</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>29.07</td>
<td>1.94</td>
<td>99</td>
<td>0.51</td>
<td>$99,770.00</td>
<td>0.36</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>5.83</td>
<td>3.77</td>
<td>118</td>
<td>0.50</td>
<td>$99,960.00</td>
<td>0.47</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.79</td>
<td>3.76</td>
<td>118</td>
<td>0.50</td>
<td>$99,600.00</td>
<td>0.47</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>2.89</td>
<td>89</td>
<td>0.49</td>
<td>$99,350.00</td>
<td>0.38</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>2.53</td>
<td>99</td>
<td>0.51</td>
<td>$99,750.00</td>
<td>0.39</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>1.41</td>
<td>149</td>
<td>0.50</td>
<td>$100,250.00</td>
<td>0.42</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>1.40</td>
<td>149</td>
<td>0.50</td>
<td>$99,950.00</td>
<td>0.42</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = 10, cJ = 50, cK = 250, R12 = 0, R22 = 0, R32 = 0, g3 = 0, P = 0.50.
Table B.1.2

Tree-level CRD design: Unequal marginal cost, starting values for n, J, K is (30, 3, 100)

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer (AUGLAG +)</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>26.82</td>
<td>2.74</td>
<td>88</td>
<td>0.50</td>
<td>$99,555.00</td>
<td>0.37</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>29.09</td>
<td>2.16</td>
<td>99</td>
<td>0.51</td>
<td>$100,198.74</td>
<td>0.38</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>6.56</td>
<td>4.74</td>
<td>104</td>
<td>0.50</td>
<td>$100,245.00</td>
<td>0.47</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.91</td>
<td>4.11</td>
<td>115</td>
<td>0.50</td>
<td>$99,858.61</td>
<td>0.48</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>2.96</td>
<td>96</td>
<td>0.50</td>
<td>$99,900.00</td>
<td>0.41</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>2.78</td>
<td>100</td>
<td>0.50</td>
<td>$100,050.00</td>
<td>0.41</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>1.98</td>
<td>122</td>
<td>0.50</td>
<td>$99,975.00</td>
<td>0.42</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>1.98</td>
<td>122</td>
<td>0.50</td>
<td>$99,975.00</td>
<td>0.42</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = c(20, 10), cJ = c(100,50), cK = (500, 250), R12 = 0, R22 = 0, R32 = 0, g3 = 0, P = 0.50.
Table B.1.3

Three-level CRD design: Equal marginal cost, starting values for n, J, K is (10, 10, 10)

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer (AUGLAG +)</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>11.47</td>
<td>22.1</td>
<td>20</td>
<td>0.50</td>
<td>$99,900.00</td>
<td>0.16</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>18.22</td>
<td>18.39</td>
<td>18</td>
<td>0.50</td>
<td>$97,910.00</td>
<td>0.15</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>7.71</td>
<td>8.33</td>
<td>58</td>
<td>0.50</td>
<td>$100,050.00</td>
<td>0.34</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.89</td>
<td>5.85</td>
<td>85</td>
<td>0.49</td>
<td>$100,230.00</td>
<td>0.42</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>19.29</td>
<td>17</td>
<td>0.47</td>
<td>$102,650.00</td>
<td>0.14</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>17.83</td>
<td>18</td>
<td>0.50</td>
<td>$100,800.00</td>
<td>0.15</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>7.17</td>
<td>42</td>
<td>0.50</td>
<td>$100,800.00</td>
<td>0.26</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>3.34</td>
<td>80</td>
<td>0.50</td>
<td>$100,100.00</td>
<td>0.36</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Note. AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = 10, cJ = 50, cK = 250, R12 = 0, R22 = 0, R32 = 0, g3 = 0, P = 0.50.
Table B.1.4

*Three-level CRD design: Unequal marginal cost, starting values for n, J, K is (10, 10, 10)*

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer (AUGLAG +)</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>13.20</td>
<td>24.32</td>
<td>22</td>
<td>0.50</td>
<td>$101,332.50</td>
<td>0.17</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>20.08</td>
<td>20.20</td>
<td>20</td>
<td>0.50</td>
<td>$98,640.00</td>
<td>0.16</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>4.75</td>
<td>7.73</td>
<td>81</td>
<td>0.49</td>
<td>$99,220.00</td>
<td>0.42</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.24</td>
<td>5.36</td>
<td>101</td>
<td>0.50</td>
<td>$99,383.42</td>
<td>0.46</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>21.26</td>
<td>19</td>
<td>0.53</td>
<td>$99,744.74</td>
<td>0.15</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>20.25</td>
<td>20</td>
<td>0.50</td>
<td>$98,625.00</td>
<td>0.16</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>5.14</td>
<td>65</td>
<td>0.49</td>
<td>$99,014.62</td>
<td>0.35</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>1.99</td>
<td>121</td>
<td>0.50</td>
<td>$99,325.62</td>
<td>0.42</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = c(20, 10), cJ = c(100,50), cK = (500, 250), R12 = 0, R22 = 0, R32 = 0, g3 = 0, P = 0.50.
Table B.1.5

Three-level CRD design: Equal marginal cost, starting values for n, J, K is (50, 50, 50)

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>4.02</td>
<td>55.33</td>
<td>12</td>
<td>0.50</td>
<td>$96,080.00</td>
<td>0.11</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>18.50</td>
<td>18.28</td>
<td>18</td>
<td>0.50</td>
<td>$98,280.00</td>
<td>0.15</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>2.26</td>
<td>21.11</td>
<td>35</td>
<td>0.51</td>
<td>$99,320.00</td>
<td>0.23</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.85</td>
<td>6.19</td>
<td>81</td>
<td>0.49</td>
<td>$99,650.00</td>
<td>0.41</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>21.47</td>
<td>15</td>
<td>0.53</td>
<td>$100,350.00</td>
<td>0.13</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>17.83</td>
<td>18</td>
<td>0.50</td>
<td>$100,800.00</td>
<td>0.15</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>3.44</td>
<td>78</td>
<td>0.50</td>
<td>$99,900.00</td>
<td>0.36</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>3.09</td>
<td>85</td>
<td>0.49</td>
<td>$100,150.00</td>
<td>0.37</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.49</td>
<td>$99,750.00</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Note. AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = 10, cJ= 50, cK = 250, R12 = 0, R22 = 0, R32 = 0, g3 = 0, P = 0.50.
### Table B.1.6

*Three-level CRD design: Unequal marginal cost, starting values for n, J, K is (50, 50, 50)*

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>12.75</td>
<td>24.50</td>
<td>22</td>
<td>0.50</td>
<td>$100,207.50</td>
<td>0.17</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>20.18</td>
<td>20.20</td>
<td>20</td>
<td>0.50</td>
<td>$98,940.00</td>
<td>0.16</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>1.5</td>
<td>60.11</td>
<td>18</td>
<td>0.50</td>
<td>$100,072.50</td>
<td>0.15</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.86</td>
<td>5.88</td>
<td>93</td>
<td>0.50</td>
<td>$99,601.72</td>
<td>0.45</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>13.45</td>
<td>29</td>
<td>0.48</td>
<td>$97,491.38</td>
<td>0.21</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>20.25</td>
<td>20</td>
<td>0.50</td>
<td>$98,625.00</td>
<td>0.16</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>4.90</td>
<td>68</td>
<td>0.50</td>
<td>$100,425.00</td>
<td>0.36</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>1.98</td>
<td>122</td>
<td>0.50</td>
<td>$100,200.00</td>
<td>0.42</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.51</td>
<td>$101,000.00</td>
<td>0.35</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = c(20, 10), cJ = c(100,50), cK = (500, 250), R12 = 0, R22 = 0, R32 = 0, g3 = 0, P = 0.50.
### B.2 Three level Cluster Random Assignment (CRA) with Randomization at Level 3

Table B.2.1

*Three-level CRA design: Equal marginal cost, starting values for n, J, K is (30, 3, 100) and for P is (0.20)*

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer (AUGLAG +)</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>27.78</td>
<td>2.58</td>
<td>81</td>
<td>0.506</td>
<td>$99,210.00</td>
<td>0.73</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>29.08</td>
<td>1.94</td>
<td>99</td>
<td>0.505</td>
<td>$99,790.00</td>
<td>0.76</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>7.19</td>
<td>3.99</td>
<td>107</td>
<td>0.495</td>
<td>$100,140.00</td>
<td>0.87</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.94</td>
<td>4.16</td>
<td>110</td>
<td>0.5</td>
<td>$100,510.00</td>
<td>0.87</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>2.89</td>
<td>89</td>
<td>0.506</td>
<td>$99,350.00</td>
<td>0.78</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>2.53</td>
<td>99</td>
<td>0.505</td>
<td>$99,750.00</td>
<td>0.80</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>1.62</td>
<td>136</td>
<td>0.5</td>
<td>$100,000.00</td>
<td>0.83</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>1.49</td>
<td>144</td>
<td>0.5</td>
<td>$100,200.00</td>
<td>0.84</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.404</td>
<td>$99,750.00</td>
<td>0.66</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.491</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.509</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.491</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = 10, cJ = 50, cK = 250, R12 = 0, R22 = 0, R32 = 0, g3 = 0.
Table B.2.2

Three-level CRA design: Unqual marginal cost, starting values for $n$, $J$, $K$ is (30, 3, 100) and for $P$ is (0.20)

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer</th>
<th>Solution for $n$</th>
<th>Solution for $J$</th>
<th>Solution for $K$</th>
<th>$P$</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>26.19</td>
<td>2.80</td>
<td>90</td>
<td>0.467</td>
<td>$99,887.33</td>
<td>0.79</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>29.73</td>
<td>2.59</td>
<td>100</td>
<td>0.31</td>
<td>$100,149.50</td>
<td>0.76</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>19.62</td>
<td>2.86</td>
<td>102</td>
<td>0.451</td>
<td>$99,740.39</td>
<td>0.83</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>14.08</td>
<td>3.47</td>
<td>104</td>
<td>0.442</td>
<td>$100,182.69</td>
<td>0.86</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>2.90</td>
<td>97</td>
<td>0.505</td>
<td>$99,942.27</td>
<td>0.82</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>2.93</td>
<td>100</td>
<td>0.45</td>
<td>$99,977.50</td>
<td>0.83</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>2</td>
<td>129</td>
<td>0.411</td>
<td>$100,100.00</td>
<td>0.84</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>2.21</td>
<td>121</td>
<td>0.421</td>
<td>$100,143.80</td>
<td>0.84</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>0.493</td>
<td>$100,000.00</td>
<td>0.75</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>96</td>
<td>0.042</td>
<td>$100,000.00</td>
<td>0.24</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>71</td>
<td>0.408</td>
<td>$100,000.00</td>
<td>0.76</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>71</td>
<td>0.408</td>
<td>$100,000.00</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Note. AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = c(20, 10), cJ = c(100,50), cK = (500, 250), R12 = 0, R22 = 0, R32 = 0, g3 = 0.
Table B.2.3

*Three-level CRA design: Equal marginal cost, starting values for n, J, K is (10, 10, 10) and for P is (0.50)*

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>12.71</td>
<td>21.20</td>
<td>20</td>
<td>0.5</td>
<td>$101,270.00</td>
<td>0.35</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>18.23</td>
<td>18.39</td>
<td>18</td>
<td>0.5</td>
<td>$97,940.00</td>
<td>0.32</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>6.35</td>
<td>7.77</td>
<td>66</td>
<td>0.5</td>
<td>$100,370.00</td>
<td>0.76</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.69</td>
<td>7.96</td>
<td>67</td>
<td>0.493</td>
<td>$100,400.00</td>
<td>0.77</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>18.41</td>
<td>17</td>
<td>0.529</td>
<td>$98,150.00</td>
<td>0.30</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>17.83</td>
<td>18</td>
<td>0.5</td>
<td>$100,800.00</td>
<td>0.32</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>9.69</td>
<td>32</td>
<td>0.5</td>
<td>$101,000.00</td>
<td>0.49</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>3.60</td>
<td>75</td>
<td>0.493</td>
<td>$99,750.00</td>
<td>0.75</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.544</td>
<td>$99,750.00</td>
<td>0.67</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.491</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.509</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.491</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented Lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = 10, cJ = 50, cK = 250, R12 = 0, R22 = 0, R32 = 0, g3 = 0.
Table B.2.4

*Three-level CRA design: Unequal marginal cost, starting values for n, J, K is (10, 10, 10) and for P is (0.50)*

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer (AUGLAG +)</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>12.65</td>
<td>21.14</td>
<td>21</td>
<td>0.81</td>
<td>$100,473.81</td>
<td>0.25</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>18.27</td>
<td>18.47</td>
<td>19</td>
<td>0.895</td>
<td>$103,007.37</td>
<td>0.16</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>5.58</td>
<td>13.52</td>
<td>52</td>
<td>0.462</td>
<td>$99,055.77</td>
<td>0.71</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.43</td>
<td>10.15</td>
<td>66</td>
<td>0.47</td>
<td>$100,233.33</td>
<td>0.78</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>21.39</td>
<td>18</td>
<td>0.611</td>
<td>$100,291.67</td>
<td>0.31</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>17.94</td>
<td>18</td>
<td>0.889</td>
<td>$100,016.67</td>
<td>0.15</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>10.62</td>
<td>37</td>
<td>0.459</td>
<td>$99,535.14</td>
<td>0.56</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>6.79</td>
<td>53</td>
<td>0.491</td>
<td>$100,240.57</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>51</td>
<td>0.941</td>
<td>$99,000.00</td>
<td>0.19</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>53</td>
<td>0.887</td>
<td>$100,000.00</td>
<td>0.31</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>71</td>
<td>0.408</td>
<td>$100,000.00</td>
<td>0.76</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>71</td>
<td>0.408</td>
<td>$100,000.00</td>
<td>0.76</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = c(20, 10), cJ = c(100, 50), cK = (500, 250), R12 = 0, R22 = 0, R32 = 0, g3 = 0.
Table B.2.6

*Three-level CRA design: Equal marginal cost, starting values for n, J, K is (50, 50, 50) and for P is (0.70)*

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer (AUGLAG +)</th>
<th>Solution for n</th>
<th>Solution for J</th>
<th>Solution for K</th>
<th>P</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>1.06</td>
<td>33.20</td>
<td>25</td>
<td>0.52</td>
<td>$98,080.00</td>
<td>0.39</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>18.51</td>
<td>18.28</td>
<td>18</td>
<td>0.5</td>
<td>$98,300.00</td>
<td>0.32</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>2.07</td>
<td>24.68</td>
<td>31</td>
<td>0.516</td>
<td>$100,100.00</td>
<td>0.49</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>5.62</td>
<td>8.20</td>
<td>65</td>
<td>0.508</td>
<td>$99,490.00</td>
<td>0.76</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>27.83</td>
<td>12</td>
<td>0.5</td>
<td>$103,200.00</td>
<td>0.22</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>17.83</td>
<td>18</td>
<td>0.5</td>
<td>$100,800.00</td>
<td>0.32</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>4.95</td>
<td>58</td>
<td>0.5</td>
<td>$100,600.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>4.49</td>
<td>63</td>
<td>0.492</td>
<td>$100,650.00</td>
<td>0.71</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.491</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.509</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.491</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.509</td>
<td>$99,750.00</td>
<td>0.68</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: MDES = 0.25, rho2 = 0.20, rho3 = 0.10, cn = 10, cJ= 50, cK = 250, R12 = 0, R22 = 0, R32 = 0, g3 = 0.
Table B.2.5

Three-level CRA design: Unequal marginal cost, starting values for $n, J, K$ is (50, 50, 50) and for $P$ is (0.70)

<table>
<thead>
<tr>
<th>Sample size constraints</th>
<th>Optimizer</th>
<th>Solution for $n$</th>
<th>Solution for $J$</th>
<th>Solution for $K$</th>
<th>$P$</th>
<th>Total cost ($100,000.00)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>COBYLA</td>
<td>7.54</td>
<td>54.07</td>
<td>15</td>
<td>0.333</td>
<td>$99,813.33$</td>
<td>0.26</td>
</tr>
<tr>
<td>None</td>
<td>MMA</td>
<td>22.34</td>
<td>22.32</td>
<td>22</td>
<td>0.182</td>
<td>$100,336.36$</td>
<td>0.25</td>
</tr>
<tr>
<td>None</td>
<td>SLSQP</td>
<td>3.75</td>
<td>14.49</td>
<td>55</td>
<td>0.455</td>
<td>$99,694.55$</td>
<td>0.72</td>
</tr>
<tr>
<td>None</td>
<td>LBFGS</td>
<td>23.31</td>
<td>10.23</td>
<td>35</td>
<td>0.486</td>
<td>$101,585.71$</td>
<td>0.54</td>
</tr>
<tr>
<td>n</td>
<td>COBYLA</td>
<td>20</td>
<td>28.93</td>
<td>15</td>
<td>0.467</td>
<td>$100,980.00$</td>
<td>0.28</td>
</tr>
<tr>
<td>n</td>
<td>MMA</td>
<td>20</td>
<td>24.25</td>
<td>24</td>
<td>0.042</td>
<td>$97,187.50$</td>
<td>0.11</td>
</tr>
<tr>
<td>n</td>
<td>SLSQP</td>
<td>20</td>
<td>9.68</td>
<td>41</td>
<td>0.439</td>
<td>$100,443.90$</td>
<td>0.59</td>
</tr>
<tr>
<td>n</td>
<td>LBFGS</td>
<td>20</td>
<td>10.72</td>
<td>36</td>
<td>0.472</td>
<td>$98,491.67$</td>
<td>0.55</td>
</tr>
<tr>
<td>n and J</td>
<td>COBYLA</td>
<td>20</td>
<td>5</td>
<td>57</td>
<td>0.754</td>
<td>$100,000.00$</td>
<td>0.55</td>
</tr>
<tr>
<td>n and J</td>
<td>MMA</td>
<td>20</td>
<td>5</td>
<td>52</td>
<td>0.923</td>
<td>$100,000.00$</td>
<td>0.23</td>
</tr>
<tr>
<td>n and J</td>
<td>SLSQP</td>
<td>20</td>
<td>5</td>
<td>71</td>
<td>0.408</td>
<td>$100,000.00$</td>
<td>0.76</td>
</tr>
<tr>
<td>n and J</td>
<td>LBFGS</td>
<td>20</td>
<td>5</td>
<td>71</td>
<td>0.408</td>
<td>$100,000.00$</td>
<td>0.76</td>
</tr>
</tbody>
</table>

*Note.* AUGLAG: Augmented lagrangian. COBYLA: Constrained optimization by linear approximations. MMA: Method of moving asymptotes. SLSQP: Sequential least-squares quadratic programming. LBFGS: Low storage BFGS. Other parameters used in PowerUpR: $MDES = 0.25$, $rh02 = 0.20$, $rho3 = 0.10$, $cn = c(20, 10)$, $cJ = c(100, 50)$, $cK = (500, 250)$, $R12 = 0$, $R22 = 0$, $R32 = 0$, $g3 = 0$. 

205
Vita

Metin Bulus was born in 1982 in Turkey. He received a Bachelor of Education degree on Science Education from Dokuz Eylul University in Izmir, Turkey; a Master of Science degree on Statistics, Measurement, Assessment and Research Technology from University of Pennsylvania, Philadelphia, USA. He was admitted to the University of Missouri in Columbia, USA in 2012 for Doctor of Philosophy Degree on Statistics, Measurement, and Evaluation in Education.